University of Alberta

Development of Enabling Sample Preparation Methods for Proteome Analysis by Mass Spectrometry

by

Alan Austin Doucette ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Chemistry

Edmonton Alberta

Spring 2002

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-68566-7

Canada

University of Alberta

Library Release Form

Name of Author: Alan Austin Doucette

Title of Thesis: Development of Enabling Sample Preparation Methods for Proteome Analysis by Mass Spectrometry

Degree: Doctor of Philosophy

Year this Degree Granted: 2002

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Alan Doucette

Box 112, R.R.#2 Tusket
Yarmouth Co., Nova Scotia
B0W 3M0

April 16, 2002

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Development of Enabling Sample Preparation Methods for Proteome Analysis by Mass Spectrometry submitted by Alan Doucette in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Liang Li
Dr. L. Li, Professor of Chemistry

Frederick F. Cantwell
Dr. F. F. Cantwell, Professor of Chemistry

D J Harrison
Dr. D. J. Harrison, Professor of Chemistry

Roderick E. Wasylishen
Dr. R. E. Wasylishen, Professor of Chemistry

Dr. A. Shaw, Professor of Oncology

Dr. R. W. Olafson, Professor of Biochemistry
and Microbiology, University of Victoria

Approved on April 15, 2002

for my wife, Deanna

Abstract

Mass Spectrometry (MS) currently plays a central role in the identification and characterization of protein samples. The associated techniques related to the preparation of samples prior to MS analysis will often dictate the ultimate applicability, and success rate of protein characterization. The work presented in this thesis is devoted to the development of novel protocols for protein sample preparation for subsequent MS analysis. These protocols combine two main aspects of protein sample preparation: methods for enzymatic degradation of proteins, and methods for sample cleanup, separation, and enrichment. The described protocols are directed at matrix-assisted laser desorption/ionization MS.

One protocol presented in this thesis is related to the application of a two-layer MALDI (matrix-assisted laser desorption/ionization) sample deposition method for protein samples containing a high concentration of sodium dodecyl sulfate. Several variables related to MALDI sample preparation are explored in this work, demonstrating that strong MALDI signals can be obtained with the two-layer method. A second protocol, described in this thesis, is related to the digestion of proteins following adsorption of the sample onto a hydrophobic chromatographic resin. The procedure is shown to be an effective method for preconcentration, as well as sample cleanup of dilute, contaminated protein samples prior to MS analysis. The method is demonstrated with standard proteins at concentrations down to 10 nM. Fundamental issues related to the digestion of adsorbed proteins are also explored in this work. A procedure is also described that is based on dual enzymatic digestion of proteins, as a means of improving the confidence in protein identification. The method is further demonstrated with real

world samples of bacterial proteins as well as a hydrophobic membrane protein. Finally, a method is presented for the enrichment of phosphoproteins from a mixture of non-phosphorylated compounds using immobilized metal affinity chromatography. This simple procedure is demonstrated with the phosphoproteins, α-casein.

The methods described in this work improve both the success and the applicability of mass spectrometry as a tool for protein identification and characterization.

Acknowledgements

I would first like to thank my supervisor, Dr. Liang Li, for his advice, support, and guidance over the course of my research at the University of Alberta. His ideas, encouragements, suggestions, and enthusiasm were always appreciated, and I feel that I have learned a significant amount from him.

I would also like to thank the many members of Dr. Li's research group that I have worked with throughout my research. In particular, I would like to thank David Craft for his contributions. In particular, his ideas, helpful discussions, and collaborative efforts in the lab lead to the development of the work described in Chapter 3 and 4 of this thesis. I am also grateful to have worked with Nancy Zhang, and would like to thank her for the enormous contributions she has made relating to the work presented in Chapter 2 of this thesis. I would also like to thank Zhengping Wang for her collaborative work and helpful discussions, as well as Bernd Keller for his help over the course of my research.

I would like to thank the personnel in the machine shop and the electronic shop of the chemistry department. I would also like to acknowledge NSERC, the Department of Chemistry, and the University of Alberta for providing me with financial support while completing my graduate degree.

Finally, I would like to thank my wife, Deanna, for providing me with her support, attention, acceptance, and most of all, patience, over the course of my studies at the University of Alberta. I realize just how big of a sacrifice she has made while I completed my degree, and I will be forever grateful for that.

Table of Contents

Chapter 1 Introduction: Protein Analysis by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry 1

1.1 The Demands of Proteome Analysis..2

1.1.1 Proteomics Defined..2

1.1.2 Separation and Characterization in Proteomics.........................2

1.2 Principles of MALDI-TOF-MS...4

1.2.1 The MALDI ionization process...4

1.2.2 The Time-of-Flight Mass Analyzer..9

1.2.3 Methods for Improving Mass Resolution in the TOF Analyzer......11

1.2.3.1 Time-Lag-Focusing..12

1.2.3.2 Reflectron TOF Instruments.................................13

1.2.4 Tandem TOF-MS instruments and CID spectra........................15

1.3 Protein Analysis by Mass Spectrometry....................................16

1.3.1 Characterization on the Basis of Intact Molecular Weight............16

1.3.2 Methods for Further Protein Characterization..........................16

1.3.2.1 Peptide Mass Fingerprinting..................................17

1.3.2.2 Tandem MS fragmentation of Proteins......................20

1.3.2.2.1 MS Fragmentation via Post Source Decay........20

1.3.2.2.2 Fragmentation via Collisional Induced Dissociation..21

1.3.2.3 Post-Translational Modification analysis by MS............24

1.4 Problems and Considerations in Sample Preparation.....................25

1.4.1 Contaminants and protein mixtures......................................25

1.4.2 Protein Concentration and Sensitivity Issues.......................... 26
1.4.3 Protein Degradation via Enzymatic Digestion.......................... 27
1.4.4 Methods to Improve MALDI Signals in Protein Analysis.......... 28

1.5 Literature Cited.. 30

Chapter 2 Two-Layer Sample Preparation Method for MALDI-Mass Spectrometric Analysis of Protein and Peptide Samples Containing Sodium Dodecyl Sulfate 36

2.1 Introduction.. 36

2.2 Experimental... 38
2.2.1 Chemicals and Reagents.. 38
2.2.2 MALDI Sample Preparation 39
2.2.3 Bacteria Extraction.. 40
2.2.4 Enzyme Digests... 40
2.2.5 Instrumentation.. 40

2.3 Results and Discussion.. 41
2.3.1 Effect of On-Probe Sample Washing.................................. 41
2.3.2 Effect of Matrix... 44
2.3.3 Effect of Solvent and Sample Preparation Method.................. 47
2.3.4 Effect of Protein Concentration... 50
2.3.5 Effect of Protein to Matrix Ratio....................................... 52
2.3.6 Blank Solutions and SDS Adduct Formation.......................... 55
2.3.7 Hydrophobic Proteins... 59
2.3.8 Bacterial Protein Extraction.. 61
2.3.9 Protein Digestion... 63

2.4 Conclusions 68

2.5 Literature Cited 69

Chapter 3 Protein Concentration and Enzyme Digestion on Micro-Beads for MALDI-TOF Peptide Mass Mapping of Proteins from Dilute Solutions 72

3.1 Introduction 72

3.2 Experimental 75
- 3.2.1 Chemicals and Reagents 75
- 3.2.2 Protein Adsorption Onto Beads 76
- 3.2.3 Enzyme Digestion of Surface-Bound Proteins 76
- 3.2.4 Enzyme Digests of Protein in Solution 77
- 3.2.5 MALDI Sample Preparation 77
- 3.2.6 Instrumentation 78

3.3 Results and Discussion 78
- 3.3.1 POROS R2 Beads 79
- 3.3.2 Optimization of Trypsin Digestion 80
- 3.3.3 Sample Deposition and Detection by MALDI-TOF-MS 80
- 3.3.4 Mechanism of Peptide Desorption From MALDI Target 84
- 3.3.5 Comparative Analysis of Uncontaminated Samples 84
- 3.3.6 Concentration Limit 88
- 3.3.7 Analysis of Samples Contaminated with NaCl and Urea 90
- 3.3.8 Analysis of Samples Contaminated with SDS 95
- 3.3.9 On-Probe Washing *vs.* Bead Washing 99
- 3.3.10 Resolution and Accuracy of MALDI Analysis 99

3.4 Conclusions...100

3.5 Literature Cited..101

Chapter 4 Investigation of the Effects of Surface-Type on the Digestion of Surface Bound Proteins in Micro-Columns 104

4.1 Introduction...104

4.2 Experimental...106
- 4.2.1 Materials and Reagents.. 106
- 4.2.2 Protein Digestion in Micro-Column.................................. 107
- 4.2.3 Digestion of Protein Adsorbed on Micro-Beads...................107
- 4.2.4 Solution Digestion of Proteins... 108
- 4.2.5 HPLC Separation and Fractionation.................................. 108
- 4.2.6 MALDI Sample Preparation... 109
- 4.2.7 Mass Spectrometry.. 109

4.3 Results and Discussion... 110
- 4.3.1 In-Column Digestion.. 110
- 4.3.2 Detailed Examination of Surface Effects on Digestion Efficiency...112
- 4.3.3 In-Column Digestion of Protein on Various Supports................ 113
- 4.3.4 HPLC Separation with UV Detection of Peptide Fragments......... 118
- 4.3.5 MS Detection of Solution and Surface-Bound Protein Digests...... 121
- 4.3.6 Effect of Bead Construction and Pore Size on Digestion............ 131

4.4 Conclusions...132

4.5 Literature Cited..133

Chapter 5 Investigation of the Applicability of a Sequential Digestion Protocol Using Trypsin and Leucine Aminopeptidase M for Protein Identification by MALDI-TOF Mass Spectrometry

.... 134

5.1 Introduction.. 134

5.2 Experimental.. 136

5.2.1 Materials and Reagents.. 136

5.2.2 Enzyme Digests.. 137

5.2.3 SDS-PAGE and In-Gel Digests.. 138

5.2.4 MALDI Sample Preparation.. 139

5.2.5 Instrumentation.. 140

5.2.6 Interpretation of Spectra.. 140

5.3 Results and Discussion.. 141

5.3.1 Sequential Digestion.. 141

5.3.2 In-Gel Digestion.. 150

5.3.3 Sensitivity.. 153

5.3.4 *E. coli* Fractions.. 157

5.3.5 CD9 Sample.. 163

5.3.6 MALDI MS/MS Fragmentation.. 167

5.4 Conclusions.. 172

5.5 Literature Cited.. 173

Chapter 6 Enrichment of Phosphorylated Proteins From a Mixture with Immobilized Metal Affinity Chromatography 176

6.1 Introduction........ 176

6.2 Experimental........ 179

6.2.1 Materials and Reagents........ 179

6.2.2 IMAC Column Preparation........ 180

6.2.3 Sample Loading, Washing, and Elution from IMAC Columns........ 180

6.2.4 Enzyme Digestions........ 180

6.2.5 Gel Separation, Staining, and Imaging........ 181

6.2.6 MALDI Sample Preparation........ 181

6.2.7 Instrumentation........ 182

6.3 Results and Discussion........ 182

6.3.1 Phosphopeptide Enrichment from Cow's Milk........ 182

6.3.2 Enrichment of Phosphoproteins by IMAC Purification........ 187

6.4 Conclusions........ 195

6.5 Literature Cited........ 196

Chapter 7 Conclusions and Future Work........ 198

List of Tables

Table 2.1 Signal to noise ratios (S/N) for 15 μM BSA samples prepared with various concentrations of SDS at different protein to matrix ratios. 53

Table 4.1 Tryptic peptides detected from the digests of cytochrome c bound to various surfaces. 116

Table 4.2 Tryptic peptides detected from the digests of BSA bound to various surfaces. 117

Table 5.1 Observed peptide fragment masses resulting from the trypsin digestion of eight protein samples in solution, each at a concentration of 0.09 μg/μl. 142

Table 5.2 Peptide fragment products of LAP digestion following trypsin digestion of protein samples. Also indicated are the times, in minutes, when the peaks appear in the MALDI spectra, and the identity of the amino acid(s) cleaved. 146

Table 5.3 Peptide fragment products of LAP digestion following trypsin digestion of 1 μg protein samples that were extracted from polyacrylamide gels. Also indicated are the times, in minutes, when the peaks appear in the MALDI spectra, and the identity of the amino acid(s) cleaved. 152

Table 5.4 Peptide fragment products of LAP digestion following trypsin digestion of 100 ng protein samples that were extracted from polyacrylamide gels. Also indicated are the times, in minutes, when the peaks appear in the spectra, and the identity of the amino acid(s) cleaved. 155

Chart 6.1 Amino acid sequence of the peptide with an observed mass of 1607.7 Da, as determined by *de novo* sequencing. 186

List of Figures

Figure 1.1	(A) MALDI-MS spectrum obtained from a sample consisting of 0.1 μg/μL cytochrome c. The sample was deposited using a two-layer deposition method with HCCA as matrix. Labeled are the singly and multiply-protonated molecular ions of cytochrome c. (B) MALDI-MS spectrum obtained from a mixture of peptides resulting from the tryptic digestion of α-casein. The sample was also deposited using a two-layer method with HCCA as matrix. 8
Figure 1.2	Schematic diagram of a reflectron time of flight mass spectrometer. Shown are the ion source region, ion mirror, and detector, along with the flight path of an ion in the instrument. 14
Figure 1.3	Nomenclature system for fragment ions generated by collisional induced fragmentation of peptides. 22
Figure 2.1	MALDI spectra obtained from samples of 500 nM BSA prepared in (A) 0% SDS (S/N=23), (B) 0.05% SDS (S/N=3), (C) 1% SDS (S/N=0), without on-probe washing; and in (D) 0% SDS (S/N=8), (E) 0.05% SDS (S/N=3), (F) 1% SDS (S/N=12), with the addition of on-probe washing. (A) & (D), (B) & (E), and (C) & (F) were from the same sample spot. The indicated signal to noise ratios are for the singly protonated ion of BSA: $[M+H]^+$. 43
Figure 2.2	Average signal to noise ratios (S/N) for the molecular ion signals calculated from the MALDI spectra of 1 μM cytochrome c at various concentrations of SDS. The matrices used for the analysis are (A) sinapinic acid, (B) HCCA, (C) HABA, (D) DHB. Error bars represent the standard deviations of S/N from 9 spectra recorded from 3 independent sample spots (A, C, D) or 3 independent spectra (B). 45
Figure 2.3	Average signal to noise ratios (S/N) for the molecular ion signals calculated from the MALDI spectra of 1 μM cytochrome c at various concentrations of SDS using HCCA as matrix with different sample deposition methods and solvent systems. The dried droplet method was used with HCCA saturated in (A) 50% acetonitrile/water and (B) 40% methanol/water. The two-layer method was used with the second layer HCCA saturated in (C) 50% acetonitrile/water and (D) 40% methanol/water. Error bars represent the standard deviations of S/N from 3 independent spectra. 48

Figure 2.4 Average signal to noise ratios (S/N) for the molecular ion signals obtained from the MALDI spectra of BSA at various concentrations of SDS. (A) 500 nM, (B) 1.5 µM, (C) 5 µM, or (D) 15 µM BSA was mixed with the second-layer matrix solution in a ratio of 1:4, and (E) 15 µM BSA was mixed in a ratio of 1:120. The second-layer matrix solution consisted of 40% methanol/water HCCA matrix. Error bars represent the standard deviations of S/N from 3 independent spectra. 51

Figure 2.5 'Blank' MALDI spectra obtained from solutions of 1% SDS with (A) HABA, (B) DHB, and (C) HCCA as matrix. Representative peaks with mass differences of 288.4 Da, or 22 Da, are labeled in the figure. 57

Figure 2.6 (A) Shown is a total MS scan of a 'blank' MALDI spectrum of a solution of 1% SDS with HCCA as matrix, recorded on a Q-star instrument. The MS/MS spectrum displayed in (B) was recorded by fragmentation of the ion from (A) with mass 1463 Da, and the MS/MS spectrum in (C) is from fragmentation of the ion with mass 887 Da. 58

Figure 2.7 MALDI spectra obtained from samples of 0.25 µg/µL bacteriorhodopsin with (A) 0% SDS, (B) 0.1% SDS and (C) 1% SDS using HCCA matrix and a two-layer deposition method. The labeled peak I indicates the bacteriorhodopsin molecular ion signal and the peak labeled as II is likely from an unprocessed precursor of bacteriorhodopsin. 60

Figure 2.8 MALDI spectra obtained from sequential extractions of an *E. coli* sample: (A) first extract with 0.1% aqueous TFA, (B) second extract with 0.1% aqueous TFA, and (C) third extract with 0.2% SDS in 0.1% TFA. 62

Figure 2.9 MALDI spectra obtained from digestion of 1 µg/µL ubiquitin (A) without adding SDS, (B) with 0.1% SDS added after digestion and (C) with 1% SDS added after digestion. The peptide fragments are labeled according to the amino acid sequence coverage of ubiquitin. 65

Figure 2.10 MALDI spectra obtained from digestion of 1 µg/µL ubiquitin (A) without adding SDS, (B) with 0.1% SDS added prior to digestion and (C) with 1% SDS added prior to digestion. The peptide fragments are labeled according to the amino acid sequence coverage of ubiquitin. 66

Figure 3.1 Distributions of micro-beads on a MALDI probe tip. About 20 μg of POROS R2 beads (20 μm diameter) were spotted on the probe tip following adsorption and digestion of a protein on the beads. Visible in the pictures are the R2 beads, shown as the large spheres, as well as the matrix crystals, shown as smaller dots. (A) The matrix solution was allowed to dry under a stream of air, which distributes the beads more evenly over the target. (B) The sample was undisturbed as the matrix solution evaporated. This caused the beads to draw together, forming non-uniform multiple layers. 83

Figure 3.2 MALDI spectra obtained from the digestion of (A) 500 nM lysozyme in solution, (B) 100 nM lysozyme in solution and (C) 100 nM lysozyme that was concentrated and digested on R2 beads. 86

Figure 3.3 MALDI spectrum showing the trypsin digestion of 10 nM cytochrome c that was concentrated on R2 beads. 500 ml of sample were used and the digestion time was 30 minutes. A total of 29 peaks were assigned to peptide fragments of cytochrome c, resulting in 97% sequence coverage. 90

Figure 3.4 MALDI spectra obtained from the free solution digestion of 500 nM BSA in (A) pure water, (B) 2 M NaCl and (C) 2 M urea. An on-probe washing procedure was incorporated for the analysis. 92

Figure 3.5 MALDI spectra obtained from micro-bead preconcentration and digestion of 100-nM BSA in (A) pure water, (B) 2 M NaCl and (C) 2 M urea. ...94

Figure 3.6 MALDI spectra obtained from the free solution digestion of 500-nM cytochrome c in (A) pure water, (B) 0.01% SDS and (C) 0.02% SDS. ..96

Figure 3.7 MALDI spectra obtained from micro-bead preconcentration and digestion of 100-nM cytochrome c in (A) pure water, (B) 0.01% SDS and (C) 0.02% SDS. 98

Figure 4.1 MALDI spectra showing the products of the digestion of 100 nM cytochrome c in-column that was packed with R2 beads for a (A) clean solution, (B) solution containing 2 M urea, 29 peaks were detected, giving 100% coverage, (C) solution containing 2 M NaCl, 27 peaks were detected, giving 97% coverage, and (D) with 0.01% SDS, 27 peaks were detected, giving 97% coverage. C = peptide fragments of cytochrome c, T = peptide fragments resulting from trypsin autolysis. 111

Figure 4.2 MALDI spectra showing the in-column digestion of 100 nM cytochrome c, where the column was packed with (A) Vydac C_{18}, (B) Vydac C_8, (C) Vydac C_4, and (D) Poros R2 beads. In the spectra, the label "C" refers to peptide fragments of cytochrome c, and "T" refers to peptide fragments resulting from trypsin autolysis. 114

Figure 4.3 UV chromatograms for the HPLC separations of peptide fragments from the digestion of cytochrome c: (A) in solution, (B) on C_{18} silica beads, and (C) on polymeric R2 beads. 120

Figure 4.4 Tryptic peptides detected from the digest of cytochrome c in solution by using MALDI and ESI-MS. The fragments observed exclusively in the direct MALDI analysis are indicated with a light gray bar, while those exclusive to LC/offline MALDI are indicated with a dark gray bar. Fragments commonly observed in each MALDI analysis method are indicated with a black bar. The number of missed cleavage sites per peptide are indicated in the figure. 123

Figure 4.5 Tryptic peptides detected from the digest of cytochrome c bound to the C_{18} surface by using MALDI and ESI-MS. The fragments observed exclusively in the direct MALDI analysis are indicated with a light gray bar, while those exclusive to LC/offline MALDI are indicated with a dark gray bar. Fragments commonly observed in each MALDI analysis method are indicated with a black bar. The number of missed cleavage sites per peptide are indicated in the figure. 124

Figure 4.6 Tryptic peptides detected from the digest of cytochrome c bound to the R2 surface by using MALDI and ESI-MS. The fragments observed exclusively in the direct MALDI analysis are indicated with a light gray bar, while those exclusive to LC/offline MALDI are indicated with a dark gray bar. Fragments commonly observed in each MALDI analysis method are indicated with a black bar. The number of missed cleavage sites per peptide are indicated in the figure. 125

Figure 5.1 (A) MALDI mass spectrum of tryptic peptides from a digestion of 1 μg cytochrome c. The peaks labeled with an asterisk correspond to peptide fragments of cytochrome c. The tryptic peptides were then further digested with leucine aminopeptidase M (LAP) and MALDI mass spectra were recorded at various LAP digestion times: (B) 2 minutes, (C) 10 minutes, (D) 30 minutes, and (E) 60 minutes. The arrows indicate the new peaks that appear as a result of LAP digestion, their parent fragment, as well as the amino acid(s) cleaved. 143

Figure 5.2 (A) MALDI mass spectrum of tryptic peptides from a digestion of 1 μg of β-lactoglobulin. The peaks labeled with an asterisk correspond to peptide fragments of β-lactoglobulin. The tryptic peptides were then further digested with leucine aminopeptidase M (LAP) and MALDI mass spectra were recorded at various LAP digestion times: (B) 2 minutes, (C) 10 minutes, (D) 17 minutes, and (E) 45 minutes. 148

Figure 5.3 (A) MALDI mass spectrum obtained from trypsin in-gel digestion of 100 ng of lactoferrin with the lactoferrin peptides labeled with an asterisk. (B) MALDI mass spectrum obtained after 2 minutes LAP digestion of the tryptic peptides. (C) MALDI mass spectrum obtained after 10 minutes LAP digestion of the tryptic peptides. 156

Figure 5.4 MALDI mass spectrum of a HPLC fraction from *E. coli* cell extract, showing an intense peak at mass 7870 Da. An expanded view of this peak is shown in the insert. 158

Figure 5.5 (A) MALDI mass spectrum obtained from trypsin digestion of the HPLC fraction shown in Figure 5.4. (B) MALDI mass spectrum recorded after 2 minutes LAP digestion of the tryptic peptides. (C) MALDI mass spectrum recorded after 10 minutes LAP digestion. 159

Figure 5.6 MALDI spectrum of a HPLC fraction from *E. coli* cell extract, showing two main protein components. An expanded view of the singly charged molecular ion peaks is shown in the insert. 161

Figure 5.7 (A) MALDI mass spectrum obtained from trypsin digestion of the HPLC fraction shown in Figure 5.6. (B) MALDI mass spectrum recorded after 2 minutes LAP digestion of the tryptic peptides. (C) MALDI mass spectrum recorded after 10 minutes LAP digestion. (D) MALDI mass spectrum recorded after 30 minutes LAP digestion. (E) MALDI mass spectrum recorded after 60 minutes LAP digestion. 162

Figure 5.8 (A) MALDI mass spectrum obtained from trypsin in-gel digestion of CD9. (B) MALDI mass spectrum recorded after 10 minutes LAP digestion of the tryptic peptides. (C) MALDI mass spectrum recorded after 30 minutes LAP digestion. 166

Figure 5.9 MALDI MS/MS spectra obtained from the fragmentation of the parent ion of mass (A) 1639.9 Da, and (B) 1511.8 Da on a Q-TOF instrument. The parent ions correspond to peptides from the solution digestion of 0.1 μg/μL BSA with trypsin and LAP. The peptide with mass 1511.8 Da results from N-terminal processing of the peptide at mass 1639.9 Da, resulting in cleavage of a single lysine residue. The detected ‘y’ and ‘b’ type fragment ions are labeled in the figure. 169

Figure 5.10 (A) MALDI MS spectra obtained from the analysis of 100 ng BSA on a q-TOF instrument. A(i) displays the spectrum from a tryptic digest of the sample and A(ii) is the spectrum obtained following LAP digestion for 10 minutes. (B) MALDI MS/MS spectrum obtained by fragmentation of the peptide with mass 1479.8 Da and (B) is the MS/MS spectrum from fragmentation of the peptide at mass 1366.7 Da. The b and y fragment ions are labeled in the figure. 171

Figure 6.1 Chemical structures of (A) phosphoserine, (B) phosphothreonine, and (C) phosphotyrosine. 177

Figure 6.2 MALDI spectra recorded from the analysis of a sample of 1:50 diluted cow’s milk. (A) is the spectrum from the undigested sample, (B) is for the sample of milk that had been digested with trypsin, and (C) was obtained following IMAC enrichment of the tryptic digest. The main protein components are labeled in (A), while the peaks corresponding to phosphopeptides are labeled in (C). The peaks labeled with an asterisk in (C) represent non-phosphorylated peptides. 184

Figure 6.3 MALDI spectra obtained from the analysis of tryptic digests of protein samples following adsorption and washing on IMAC beads. (A) was obtained by loading α-casein on IMAC beads, with an acetic acid wash (pH 3.5) (B) was from loading α-casein and cytochrome c, with acetic acid wash (pH 3.5), and (C) was from loading α-casein and cytochrome c, with an NH$_4$OH wash (pH 12) prior to digestion. Peptide fragments of cytochrome c and of α-casein (α_{S1} or α_{S2}) are labeled ‘cytc’ or ‘α_{S1}’ and “α_{S2}’ respectively. 189

Figure 6.4 Gel image obtained from electrophoresis of (I) the basic extract (pH ~9.5) of an IMAC column loaded with α-casein and cytochrome c (II) a 50% acetonitrile/0.1% TFA extraction of the same column following the first extraction and (III) a standard consisting of 1 μg α-casein and 1 μg cytochrome c. 191

Figure 6.5 MALDI spectra obtained from tryptic digests of a mixture of α-casein and cytochrome c following an IMAC purification procedure, with extraction from the IMAC column with various solutions. (A) The IMAC column was extracted with an aqueous solution at pH 7.5, (B) at pH 9, (C) at pH 12, and (D) extracted with a solution of 50% acetonitrile/0.1% aqueous TFA. The peaks are labeled as in Figure 6.4. 194

Figure 7.1 Image of sample well and frit plate. 202

List of Abbreviations

Å	angstrom
ATCC	American Type Culture Collection
2D	two-dimensional
BSA	bovine serum albumin
C	celcius
CID	collisional induced dissociation
Da	daltons
DHB	2,5-dihydroxybenzoic acid
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EC	enzyme commission
ECL	enhanced chemiluminescence
ESI	electrospray ionization
FWHM	full width at half maximum
HABA	2-(4-hydroxyphenylazo)-benzoic acid
HCCA	α-cyano-4-hydroxycinnamic acid
HPLC	high pressure liquid chromatography
HRP	horseradish peroxidase
I.D.	internal diameter
IMAC	immobilized metal affinity chromatography
IR	infrared

kDa	kilodalton
KE	kinetic energy
LAP	leucine aminopeptidase
LC	liquid chromatography
M	molar
mAb	monoclonal antibody
MALDI	matrix-assisted laser desorption/ionization
μg	microgram
mg	milligram
min	minute
μL	microliter
mL	milliliter
mM	millimolar
MS	mass spectrometry
MS/MS	tandem-mass spectrometry
m/z	mass to charge ratio
Nd:YAG	neodinium/yttrium/aluminum/garnet
ng	nanogram
nM	nanomolar
nmol	nanomole
pmol	picomole
PAGE	polyacrylamide gel electrophoresis
pI	isoelectric point

ppm	parts per million
PSD	post source decay
PTM	post-translational modification
Q-TOF	quadrupole-time-of-flight
SA	sinapinic acid
SDS	sodium dodecyl sulfate
S/N	signal to noise ratio
TFA	trifluoroacetic acid
TLF	time-lag focusing
TOF	time-of-flight
UV	ultraviolet
v/v	volume to volume ratio
w/v	weight to volume ratio

Chapter 1

Introduction:

Protein Analysis by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry

The last decade has seen a revolution in biological science. In its early stages, this revolution was driven by the development of high-throughput DNA sequencing technology [1]. Now, in a post-genomic era, advancements in protein characterization are moving towards the analysis of proteins in a parallel, high-throughput fashion similar to that of genomics. Currently, mass spectrometry (MS) takes a leading role in the technology for protein analysis [2], owing to the advent of efficient ionization methods for biopolymers, namely electrospray ionization and matrix-assisted laser desorption/ionization. Recently, an explosion in growth has occurred relating to the development of new instrumentation in mass spectrometry, attempting to meet the incredible demands of protein characterization [3]. However, numerous challenges have yet to be met. One such area is associated with the preparation of problematic samples prior to MS analysis, such as dilute or contaminated protein mixtures, or the analysis of hydrophobic proteins. Difficulties associated with such samples lower the ultimate success rate of protein characterization.

In this work, current methodologies in protein analysis by mass spectrometry will be described, along with their associated limitations. This work further describes novel approaches that address these limitations and offer improvements in the analysis of protein samples by mass spectrometry.

1.1 The Demands of Proteome Analysis

1.1.1 Proteomics Defined. Wilkins first introduced the term "proteome" or "proteomics" in 1995 [4]. It denotes the full protein complement of a cell, or cell culture, expressed at a given state or time. Unlike its genome counterpart, a proteome is a dynamic system; a proteome changes in response to various conditions or external stimuli, including stages of cell or tissue growth, culture conditions, or pathological invasion of the tissue [5]. A proteome, therefore, provides a kind of snapshot view of a cell at a given point in time. Compared with the genome, a proteome is a much more complex system, owing both to the diversity of protein sequence and structure, as well as the large number of proteins expressed at a given time [5,6]. From completion of a rough draft of the human genome, it was reported that our genome contains on the order of 30,000 to 40,000 protein-coding genes [7, 8]. Other considerations, such as gene splicing and post-translational protein modifications, give rise to a complete human proteome that is expected to be even larger (perhaps up to 10 times more) than the corresponding gene number [6]. The structural properties of the proteins contained in this mixture will vary considerably. Protein size can range from less than 10 kDa, to larger proteins that are greater than 100 kDa. Also, considerable differences exist in other properties such as protein isoelectric point (pI) and hydrophobicity. To further complicate this mixture, typical protein abundances in a proteome can have a dynamic range of up to 10^6 [5].

1.1.2 Separation and Characterization in Proteomics. Owing to the complexity of a proteome mixture, its analysis requires addressing two important issues. The first lies in the preparation steps (*i.e.* the extraction, purification and separation processes). To this end, a separation of the enormous mixture into individual proteins, or at the very least, a partial separation, is often key to the success of the analysis [9,10]. Of the many techniques available for protein separation, two-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (2D-SDS-PAGE), which was developed and first applied to proteins in 1975, remains a pivotal technique [11]. This method allows a large number of proteins (up to 5000) to be simultaneously separated and displayed. The second issue relates to the characterization of the proteins themselves once a initial purification/separation stage is complete. Traditionally, proteins were analyzed in a slow, tedious fashion in which the partial or complete amino acid sequence of the protein was determined, typically through Edman degradation [12]. This sequencing process has been coupled to 2D gel-separated proteins that were electroblotted onto membranes [13,14]. In the technique, the protein is digested through enzyme proteolysis (trypsin being one of the enzymes of choice), and the resulting protein fragments are individually characterized according to their amino acid sequence by Edman degradation. However, this technique, which is now referred to as Edman microsequencing, lacks the speed, and sensitivity necessary to analyze the complete range of proteins displayed in a proteome [9]. Today, mass spectrometry has taken a leading role as a technique for rapid protein analysis. Mass spectrometry saw its initial breakthrough in protein analysis in the 1980's with the development of several new ionization techniques [15]. These techniques made it possible to introduce a variety of high molecular weight, thermally labile biopolymers into the mass analyzer with minimal fragmentation. Of the many ionization methods developed, two techniques, namely electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) have gained the most widespread use for protein analysis. A discussion of the MALDI process will be presented here, since it is the technique that has been extensively used in this work.

1.2 Principles of MALDI-TOF-MS

1.2.1 The MALDI Ionization Process. Several techniques are available to efficiently ionize analytes from a solid support, however matrix-assisted laser desorption/ionization (MALDI) is most commonly employed for protein and peptide applications with mass spectrometry. The MALDI technique saw its beginnings in 1987 when two groups independently reported variations of the MALDI technique as applied to the analysis of high molecular weight proteins [16-19]. The research group lead by Tanaka [16,17] reported a MALDI protocol using a matrix consisting of ultra fine cobalt powder suspended in glycerol. With this, increased sample volatility of high molecular weight proteins (up to 100 kDa) suspended in the matrix was achieved. This was a consequence of rapid heating through absorption of the laser energy by the fine metal powder. Although this form of MALDI technique is still in current practice [20], the MALDI technique introduced by Karas and Hillenkamp in 1987 is by far the most widely employed protocol today.

The MALDI technique of Karas and Hillenkamp [18,19] evolved as a modification of the principles of laser desorption ionization. Earlier observations showed that the UV laser energy used to desorb the sample could be significantly lowered while still achieving effective ionization if the wavelength was tuned to the resonant electronic excitation of the sample [21,22]. Lower threshold energy thus allows for ionization with a lower degree of fragmentation; a key feature in allowing for the detection of intact molecular ions by mass spectrometry. As described by Hillenkamp [23], in a pivotal experiment, Karas ran a mixture of tryptophane and alanine using laser desorption mass spectrometry. Although the laser fluence was adjusted to desorb only tryptophane, a strong signal for alanine was also observed. Desorption of tryptophane acted to carry the alanine along with it. MALDI was born! Based on these observations, Karas and Hillenkamp later incorporated small, UV absorbing molecules as a matrix for desorption of proteins [18, 19]. The principles of this technique are described below.

In a typical MALDI setup, the solid matrix molecules are first dissolved in an appropriate solvent system and then mixed with a low concentration of analyte molecules. The mole ratio of matrix to analyte has a wide range of between 500:1 and 10,000:1 [24]. This matrix/analyte solution is deposited on a metal target and allowed to coprecipitate as the solvent evaporates. Several strategies for sample deposition have been presented in the literature to provide the best crystallization of analyte and matrix. These include dried-droplet [19], vacuum drying [25], crushed-crystal [26], slow crystal growing [27], active film [28,29], pneumatic spray [30], electrospray [31], fast solvent evaporation [32], sandwich [33,34], and two-layer methods [35]. Following sample deposition and crystallization, the resulting crystal (which has been termed a "solid solution") incorporates analyte molecules that are evenly dispersed or solvated between the matrix molecules (herein referred to as co-crystallized) [36]. The dispersion of analyte molecules has the benefit of reducing intermolecular interactions between analyte molecules, which ultimately aids in its volatility. The main purpose of the matrix, however, is to absorb photon energy. In doing so, the matrix protects the analyte molecules from direct adsorption of the photon energy, which would otherwise lead to fragmentation of the sample [36, 37]. The MALDI sample is analyzed by irradiating the crystals with a pulsed laser. The type of laser employed is chosen according to the resonant electronic excitation energy of the matrix. Both UV lasers, including Nd:YAG and N_2 lasers, and IR lasers such as a CO_2 laser have seen various applications, with UV-MALDI experiencing the best results for protein and peptide analysis [23,37]. As the laser pulse hits the matrix crystals, the matrix absorbs the photon energy, which causes a rapid heating of the sample. This essentially results in an explosion of the first few layers of the matrix crystal, where matrix molecules are lifted from the solid support by way of a supersonic expansion. Since the analyte molecules are distributed among the matrix crystals, the resulting expansion of matrix also entrains analyte molecules into the gas phase [38]. The entire process can be modeled as a phase transition from solid to a high-pressure fluid [38].

The process of ionization of the analyte molecules also occurs as a consequence of irradiation of the sample. A single theory to explain the ionization process has not yet been accepted; several researchers have attempted to present models for the ionization process [36,38-41], and the topic has been recently reviewed [42]. It is believed that the matrix contributes to sample ionization by acting as a proton donor to the analyte molecules. Initial ionization of the matrix can occur through various processes, such as multiphoton ionization, excited state proton transfer or thermal ionization. Ultimately, the end result of the MALDI process is the generation of predominantly singly protonated molecular ions, as well as some multiply protonated ions of the analyte species.

Figure 1.1 shows representative examples of MALDI-MS spectra obtained from samples consisting of (A) cytochrome c, and (B) a peptide mixture resulting from a digestion of α-casein with trypsin. The samples were prepared using a two-layer deposition method [35]. The matrix used in these preparations was alpha-cyano-4-hydroxy cinnamic acid (HCCA). The spectrum in Figure 1.1A (that of cytochrome c) clearly displays the singly protonated molecular ion $[M+H]^+$ of cytochrome c, along with multiply-protonated ion of the protein. The low mass region of the spectrum exhibits the typical low-mass background of a MALDI spectrum, which consists of matrix ions, various fragments of matrix ions, matrix clusters such as dimers and trimers, as well as adducts of the matrix molecules with metal ions such as Na^+ and K^+. Software programs have been developed that can be used to predict the mass of the matrix ions that contributes to the low mass background in the spectrum [43].

Figure 1.1B shows the power of a MALDI-MS spectrum to analyze mixtures of peptides. In this spectrum, several peptide fragments are clearly detected. These peptides were generated by an enzymatic digestion if the protein α-casein with trypsin. The α-casein was prepared by dissolving the crystalline solid, from Sigma Aldrich Canada (Oakville, ON), in water to a concentration of 0.1 μg/μL. Again, a two-layer MALDI deposition method was employed to analyze the sample, using HCCA as matrix.

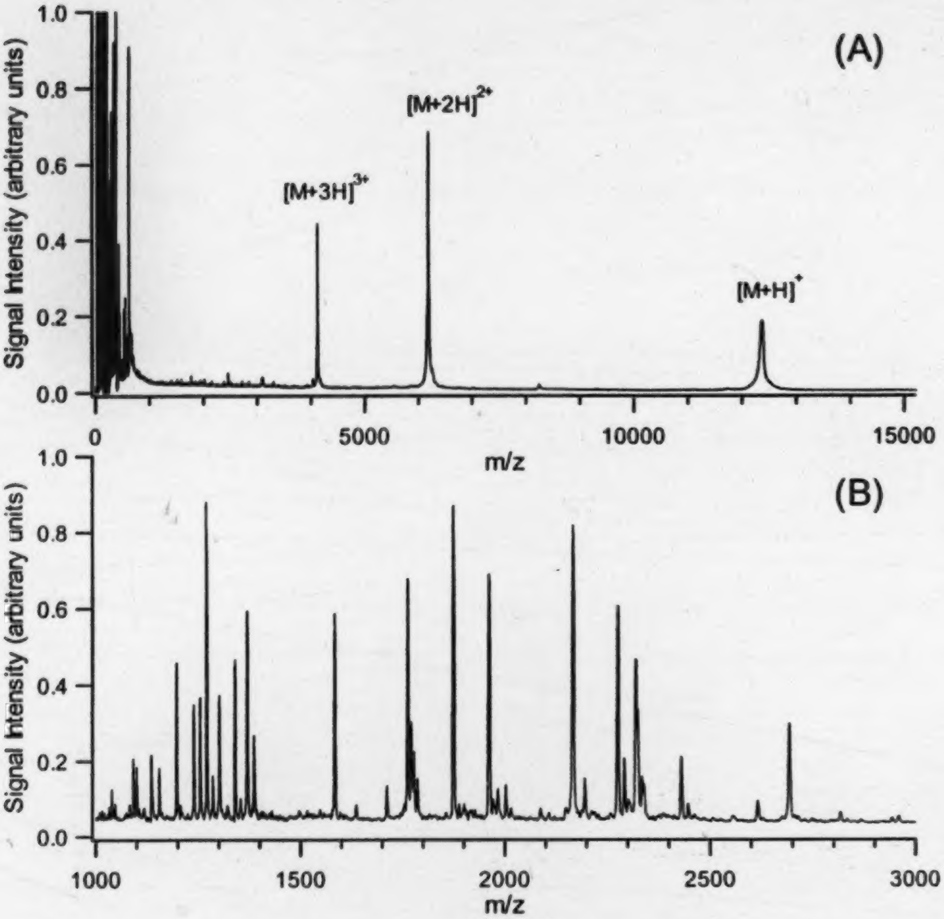

Figure 1.1 (A) MALDI-MS spectrum obtained from a sample consisting of 0.1 μg/μL cytochrome c. The sample was deposited using a two-layer deposition method with HCCA as matrix. Labeled are the singly and multiply-protonated molecular ions of cytochrome c. (B) MALDI-MS spectrum obtained from a mixture of peptides resulting from the tryptic digestion of α-casein. The sample was also deposited using a two-layer method with HCCA as matrix.

1.2.2 The Time-of-Flight Mass Analyzer. The MALDI technique generates ions from a solid support in a pulsed format. Such a system is ideally suited for coupling with a time-of-flight (TOF) mass spectrometer. In fact, the development and application of MALDI-MS contributed significantly to the current popularity of these instruments. TOF mass spectrometers are among the simplest, most inexpensive instruments for mass analysis [44]. They are characterized by high sensitivity owing to their ability to detect ions of all masses simultaneously and have high transmission efficiency compared to other mass spectrometers [45]. Also important, particularly for biological applications, is that TOF instruments have an essentially unlimited mass range.

TOF mass analyzers are far from new. In 1946, Stephens first proposed the idea of a device for the separation of ions in a vacuum tube according to their mass-based differences in their velocity as they travel through the drift tube [46]. The principles governing this technique can most easily be explained from fundamental equations.

It is known that an ion placed in an electric field, E, will be imparted kinetic energy (KE) according to the equation given as

$$KE = zeE\,, \tag{1}$$

where e is the electron charge and z is the number of charges on the ion. Also, the kinetic energy of the ion can be related to its mass, m, according to the equation below:

$$KE = \frac{1}{2} m v^2\,, \tag{2}$$

where v is the final velocity of the ion. From the first equation, we can see that all ions of equal charge placed in an identical electric field will be imparted the same kinetic energy. Also, from the second equation, we can see that once given a set magnitude of kinetic energy, the velocity of the ion will vary if the masses of these ions are different.

Therefore, if these ions are made to travel through a field free drift tube, then an expression for the time, t, required to traverse the drift tube can be found by combining the first two equations. This expression is given as

$$t = \left(\frac{m}{2zeE} \right)^{\frac{1}{2}} D, \tag{3}$$

where D is the length of the drift tube. By rearranging the above equation, we can see that the flight time of an ion in a TOF instrument is related to its mass to charge ratio according to the following equation:

$$\frac{m}{z} = 2eE\left(\frac{t}{D} \right)^{2}. \tag{4}$$

Thus, in a given experiment using a TOF analyzer, where the length of the drift tube (D) is fixed, and for a predetermined accelerating electric field (E), the mass to charge ratio of an ion can be directly determined from the flight time of that ion.

In practice, a time-of-flight instrument measures the flight time of an ion by placing an ion detector, such as a multichannel plate at the end of the flight tube, which records the ion current in relation to the time at which the ions are made to accelerate in an ion source region at the opposite end of the flight tube. Sophisticated timing electronics are required for accurate measurement of the ion flight times. Owing to the development of more advanced computer systems and digitizers, which no longer provide a bottleneck to the resolution of the instrument, a modern TOF instrument is capable of measuring the flight time of ions, and thus its mass to charge, with an extreme degree of accuracy [47]. The mass scale of a TOF analyzer is determined by calibrating the instrument with standard compounds of known mass. However, uncertainties in three other critical parameters, namely the initial temporal, spatial, and velocity distribution of the formed ions, contribute to variability in the flight time, even for those ions having identical mass to charge [48].

1.2.3 Methods for Improving Mass Resolution in the TOF Analyzer. Finite resolution in a mass analyzer contributes to an observed peak width in a mass spectrum for ions of a given mass to charge ratio. The resolution, R, in a TOF instrument is typically defined as the peak width at half height, which is also related to the flight time of the ion as follows:

$$R = \frac{m}{\Delta m} = \frac{t}{2\Delta t}. \tag{5}$$

From this equation, one can see that an improvement in resolution can be achieved by reducing the time width (Δt) of ions hitting the detector at the end of the flight tube. Thus it is required that the defined starting point be known with the highest degree of certainty. This can be most easily achieved by using a very short laser pulse to generate the ions in the shortest possible time window. The second factor contributing to decreased resolution, namely the spatial distribution of the ions, can be minimized in a TOF experiment by use of MALDI in the preparation of the sample for ionization. In MALDI, all ions are desorbed from a very thin layer of the solid support, and thus spatial distribution is minimized. Proper preparation of the MALDI sample crystals on the sample target can further enhance resolution.

The final factor that decreases resolution in a TOF analyzer, namely the initial kinetic energy distribution of the ions, has been the most difficult parameter to overcome. However, this has been achieved through the development and implementation of various compensation techniques that ultimately narrow the time gap at which ions of different kinetic energies arrive at the detector. Two such compensation techniques, the delayed extraction or time-lag-focusing technique, and the ion mirror, or reflectron time-of-flight instrument will be explained in the sections to follow.

1.2.3.1 Time-Lag Focusing. In a conventional MALDI-TOF instrument, the ions generated from the pulsed laser are immediately subjected to an electric field since a constant dc voltage is continuously applied to the ion source plate. The main problem associated with this system, which is referred to as continuous extraction, relates to the initial kinetic energy distribution of the ions [49]. In order to explain this effect more clearly, let us assume that there are only two ions in the source region, both having an identical mass to charge ratio. Also, let us assume that ion 1 possesses a greater initial kinetic energy than ion 2 (a result of the ionization process). In addition to the initial kinetic energy possessed by these ions, both ions are immediately imparted an identical amount of kinetic energy as they experience the same electric field. The magnitudes of these kinetic energies can be summed, and thus ion 1 will have a higher total kinetic energy than ion 2. The excess kinetic energy possessed by ion 1 translates into a higher velocity in the direction of the detector at the end of the flight tube. The net result is that, although these ions have identical mass to charge, ion 1 will reach the detector before ion 2. In a real system, where several ions are generated in the source region, the net effect of the initial kinetic energy distribution of ions at a given mass to charge ratio is a broadening of the ion cloud at the detector, resulting in a decrease in mass resolution.

Wiley and McLaren first described the concept of time-lag focusing (TLF) in 1955 [50]. In a TOF instrument equipped with TLF, the ions are first generated in a field free region. Due to the initial kinetic energy of the formed ions, these ions will drift away from the source plate. After a set delay time following generation of the ions (equivalent to the time of the laser pulsing in a MALDI-TOF instrument), an accelerating voltage is applied to the source (repeller) plate. Since the ions are now spatially distributed in the source region, the effect of TLF is to impart a different amount of kinetic energy to the ions. Those ions closest to the extraction plate will be imparted higher kinetic energy than the ions that drifted further away from the extraction plate, which in turn reduces the time width of the ions of given m/z as they arrive at the detector.

Again we can explain the effect assuming only two ions of the same m/z are generated in the source region, with ion 1 having higher initial kinetic energy than ion 2. After a set delay time, ion 1 will have traveled further away from the repeller plate than ion 2. The extraction voltage is then applied to the repeller plate. Since ion 1 is now further away from the repeller plate than ion 1, it will travel through a smaller electric potential drop, and thus the field imparts less kinetic energy to ion 1 than to ion 2. Therefore, ion 2 will travel faster than ion 1 and as these ions travel through the drift tube of the TOF analyzer, ion 2 will be able to catch up to ion 1. In an ideal case, the point of focusing of these two ions will coincide with the position of the detector in the TOF instrument, resulting in improved mass resolution for the detected ions. In practice, the delay time, magnitude of the extraction voltage, and shape of the extraction pulse are adjusted according to the mass of the analyte ions in order to optimize the TOF instrument resolution [51].

1.2.3.2 Reflectron TOF Instruments. The reflectron TOF instrument was developed in 1973 by Mamyrin [52]. The reflectron, otherwise known as an ion mirror, consists of a series of ring electrodes placed at the end of the drift tube. A schematic of the instrument can be seen in Figure 1.2. The voltage applied to the ring electrodes increases in potential as one moves deeper into the ion mirror. As ions enter the ion mirror, they penetrate to a point where they reach zero energy, and then are reaccelerated in the near opposite direction, with the same kinetic energy that they entered the ion mirror. For ions having higher kinetic energy, these ions will penetrate further into the ion mirror, and thus travel a longer distance than ions with lower kinetic energy. As a result, the longer flight path will compensate for those ions with higher initial kinetic energies, thus improving resolution. Resolution on the order of 37,000 (FWHM) has been recorded for proteins with modern reflectron TOF instruments [15].

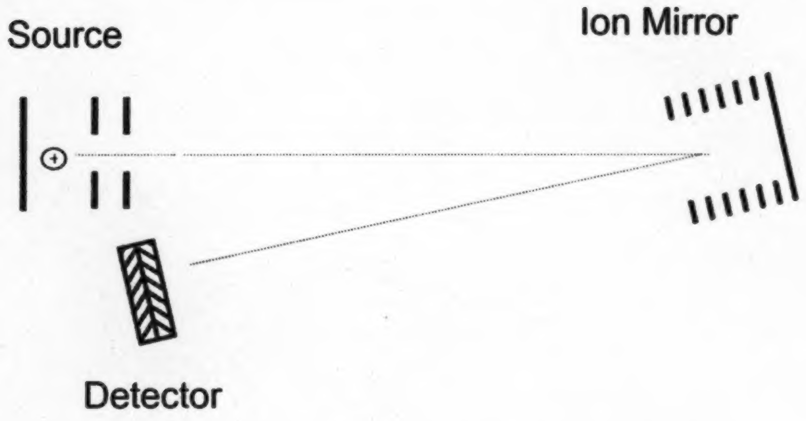

Figure 1.2 Schematic diagram of a reflectron time of flight mass spectrometer. Shown are the ion source region, ion mirror, and detector, along with the flight path of an ion in the instrument.

1.2.4 Tandem TOF-MS Instruments and CID Spectra. Time-of-flight instruments have seen a revival since the introduction of MALDI for the analysis of biopolymers. Significant advances in the development of reflectron instruments with improved ion focusing optics, as well as time-lag focusing have considerably expanded the attainable resolution with these instruments. As a result, the typical mass accuracy in modern instruments is on the order of 5 ppm [12]. However, improvements in resolution and mass accuracy are not the sole driving force in the development of new TOF-based MS instruments. An example of the new applications of TOF instrumentation is demonstrated by the development and commercialization of two new tandem time-of-flight instruments, namely the TOF-TOF [53] and the quadrupole-TOF (Q-TOF) instrument [54]. Using these tandem instruments, analyte ions can be fragmented and subsequently analyzed by the MS instrument and therefore provide structural information on the analyte.

In a TOF-TOF or Q-TOF instrument, fragmentation of the analyte is accomplished through the process of collisional induced dissociation (CID). In this process, the analyte ions are made to fragment by colliding these ions with a gas. The collisions between analyte and gas transfer kinetic energy into internal energy, which breaks chemical bonds of the analyte. The type of information that can be gained using these instruments will be discussed in section 1.3.2.2: Tandem MS Fragmentation of Proteins.

1.3 Protein Analysis by Mass Spectrometry

1.3.1 Characterization on the Basis of Intact Molecular Weight. As mentioned previously, the initial breakthrough in protein analysis by mass spectrometry came at the end of the 1980's with the development of MALDI and ESI. With these techniques, nonvolatile, thermally labile protein samples could be introduced to the mass analyzer as intact molecules. Coupled to instruments such as the time-of-flight mass spectrometer, MS instruments can now be used to accurately determine the molecular weight of extremely large biopolymers; literature accounts on the measurement of proteins with masses exceeding 1 million Daltons have been reported [55,56]. Intact molecular weight information for an unknown protein provides a first level of protein characterization, and is often sufficient to confirm the identity of a protein sample. Also, differences between the observed and predicted molecular weight of a protein can provide an insight into protein modification, such as proteolysis, or the addition of chemical moieties to the protein such as phosphorylation or carbohydrate attachments. Therefore the determined molecular weight of a protein sample may differ from its predicted value.

1.3.2 Methods for Further Protein Characterization. Protein analysis is not limited to the determination of the molecular weight of the intact molecule. Although this knowledge provides useful insight, it is somewhat limited and does not fully characterize an unknown sample. Coupled with various sample manipulation protocols, mass spectrometers can provide a wealth of additional information on the structural composition of protein samples. Specifically, MS based methods can be used to determine the amino acid sequence of a protein (referred to as its primary structure). Mass spectrometry has also been used to provide information on higher order protein structure (secondary, tertiary and quaternary structures) [12]. Thus, mass spectrometry has become an effective instrument in protein characterization.

The unique ability of mass spectrometry to characterize protein samples, particularly on a high throughput platform, became reality in the early 1990's when researchers developed platforms for protein identification that take advantage of the extensive information on protein sequence already contained within computer databases. [57-61]. A key feature of these techniques is that it was not necessary to determine the amino acid sequence of the sample protein in order to make an identification. Two fundamental methods that are based on this principle include the techniques of peptide mass mapping and amino acid sequencing via peptide fragmentation for searching protein databases. These techniques for protein analysis and identification will be discussed in more detail in the following section.

1.3.2.1 Peptide Mass Fingerprinting. A pivotal development in the field of mass spectrometry came when several research groups independently reported on a method for rapid protein identification termed peptide mass fingerprinting [57-61]. The technique takes full advantage of the remarkable speed, sensitivity, and selectivity of the MALDI or ESI techniques for mass spectrometric analysis.

In peptide mass fingerprinting, a protein sample is first degraded in a controlled fashion by using a sequence specific enzyme to digest the protein. The enzyme trypsin has been used most frequently to digest the protein, as it will selectively cleave on the carboxy-terminal side of two amino acids, namely arginine and lysine. The digestion results in the production of several peptide fragments of the protein. The masses of these peptides are then accurately determined using mass spectrometry. The obtained masses are then entered into a computer based search algorithm. This algorithm correlates the entered mass data against a databank of proteins that can be digested *in silico*. Tentative identification is made when the search algorithm finds a protein in the databank with the highest degree of peptides matching a theoretical digest of that protein. Some examples of the protein database searching programs for peptide mass mapping that are available on the Internet include PeptIdent (http://ca.expasy.org/tools/peptident.html), MS-FIT (http://prospector.ucsf.edu/ucsfhtml4.0u/msfit.htm), and the searching program Mascot (www.matrixscience.com).

Various criteria are employed by the search algorithms in correlating peptide mass information with protein database entries. In the simplest approach, such as that used by the PeptIdent searching program, the output (*i.e.* the potentially matching proteins from the database) is ranked, or scored, according to the number of peptide fragments found for a given protein entry. The highest score is assigned to the protein(s) with the highest number of peptide fragments having masses that match the input list of peptides used to search the database. The MS-FIT program involves a higher level of scoring to rank the output of the search results. Using the MOWSE (molecular weight search) search algorithm, peptide fragment masses that match to theoretical fragments of the database are weighted according to the frequency of occurrence of such fragments within a set mass range [61]. The final score in the output protein list is therefore based on a weighted sum of individual scores for the matching peptide fragments. An even higher level of statistical treatment is involved incorporated in the MASCOT search algorithm. In this case, the output protein list is scored based on the MOWSE program, however the output scores are further treated to take random matches of peptides into consideration [62]. The search result is therefore a probability-based scoring of the matching proteins from the database.

Of course, the key to the success of this technique is that the sequence of the protein analyzed must already be known and contained in a database. Fortunately, the rapid progression in genomic sequencing of various species contributes to an exponential growth in the number of entries in protein sequence databanks. Several databanks are currently accessible on the Internet. Examples of such databases include SwissProt, a protein sequence databank that has over 100,000 entries as of January 31st, 2002, and TrEMBL, a supplement to SWISS-PROT containing translations of DNA coding sequences, with almost 560,000 entries at the same time. These databases can be accessed at http://ca.expasy.org. Other databases containing protein sequence information include OWL (http://www.bioinf.man.ac.uk/dbbrowser/OWL/), NCBInr (http://www.ncbi.nlm.nih.gov/), and MSDB (ftp://ftp.ncbi.nih.gov/repository/MSDB/msdb.nam).

In peptide mass fingerprinting, the basic goal is to generate only a minimal amount of information (the mass of the peptide fragments) to extract the full amino acid sequence of the protein of interest from a computer database. In collecting the mass information, a higher degree of mass accuracy will constrain the search to fewer possible choices for matching peptides, thus improving the success rate of finding the identity of the protein [10,12]. Another important aspect of peptide mass mapping relates to the successful generation of the specific peptide fragments. To this end, the choice of enzyme used to fragment the protein is an important consideration. In some cases, the use of multiple enzymes, in tandem, or in sequence, may aid in the generation of the peptides, or can improve the confidence in identification of the protein. The use of multiple enzymes for peptide mass mapping will be discussed further in Chapter 5 of this work.

1.3.2.2 Tandem MS Fragmentation of Proteins. Often, the information gained by peptide mass mapping is not sufficient to positively identify an unknown protein sample. In particular, it is not always possible to generate a sufficient number of peptides to provide a confident identification of the unknown protein. Also, when dealing with protein mixtures, peptide mass mapping becomes difficult as peptides will be generated and detected resulting from the various proteins in the mixture. Other problems associated with peptide mass mapping include non-specific enzyme digestion of the protein, and post-translational modifications (PTM's), which lead to the generation of peptides of unknown mass. A complementary approach to peptide mass mapping is based on inducing fragmentation of the protein or peptide, and then relating these fragments directly to the amino acid sequence. Combined with other information such as the parent peptide mass, the determination of a partial amino acid sequence in the protein (a stretch of as few as 2 amino acids) will often positively identify an unknown contained within the protein sequence database [63].

1.3.2.2.1 MS Fragmentation via Post Source Decay. Protein fragmentation can be accomplished using a variety of mass spectrometric instrumentation techniques. Traditionally, amino acid sequence information was obtained using a reflectron time-of-flight mass spectrometer with a method known as post-source decay [64]. In this process, following sample ionization, some of the generated ions will fragment or breakdown as a result of energy transfer to the ions during the ionization process. Although the ion fragments are formed with identical velocities as their parent ions, the change in mass will give the fragments a unique kinetic energy. Thus as these fragment ions enter the ion mirror of the reflectron, they will penetrate the ion mirror to different extents. The result of this is that the reflectron instrument can separate the fragment ions, and these fragments can be related to compositional information on the analyte.

1.3.2.2.2 Fragmentation via Collisional Induced Dissociation. The most effective method of generating fragments is through a process known as collisionally induced dissociation (CID). This technique has been traditionally accomplished in either an ion trap or a triple quadrupole mass spectrometer. However, the recent commercialization of quadrupole-TOF and TOF-TOF instruments now allow for CID to be routinely performed in TOF instruments as well.

The process of CID involves colliding the ionized analyte ions with a relatively high pressure of gas, such as He, N_2, or Argon gas. Kinetic energy from the collision is transferred into vibrational energy, which breaks the internal bonds of the analyte ions.

In low energy CID of peptides, the peptide backbone is fragmented in a characteristic fashion. A system of nomenclature was developed for the characteristic ions that are formed [65]. The system is based on the position of bond breakage, as well as whether the charge is carried on the amino-terminal or carboxy-terminal side of the peptide. The nomenclature for fragment ions is summarized in Figure 1.3.

From Figure 1.3, it is seen that a, b, and c type fragment ions refer to the charge remaining on the amino-terminal side of the peptide, and x, y, and z type fragment ions have the charge on the carboxy-terminal side. The subscripts (1, 2, 3…) in the fragment ion nomenclature refer to the position of bond breakage in relation to the number of amino acid residues that are present on the fragment ion.

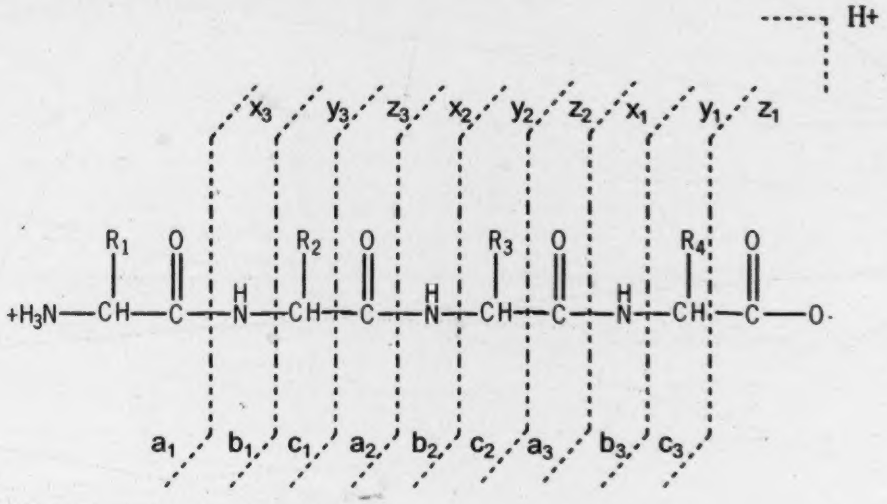

Figure 1.3 Nomenclature system for fragment ions generated by collisional induced fragmentation of peptides.

CID spectra of peptides can provide a wealth of information on the composition of the peptide. In the first case, let us assume the sequence of the given peptide is known, or suspected. Here, a computer program can be used to generate a list of all possible fragments of that peptide. An example of such a program is MS-Product, which is available from the UCSF Mass Spectrometry Facility website at http://prospector.ucsf.edu/ucsfhtml4.0u/msprod.htm. Through correlation of the obtained CID spectrum with the list of possible fragment masses of the peptide, the expected amino acid sequence of the peptide can either be confirmed, or rejected. If the sequence information is confirmed, then the expected identity of the protein is also confirmed.

In the second case, let us assume that the amino acid sequence of the peptide to be fragmented is completely unknown. In this case the fragments detected in the CID spectrum can be used as a search parameter against the protein sequence databank. The computer based search algorithm works on essentially the same principle as that of peptide mass fingerprinting. However, in this case, the use of fragments that relate to amino acid sequence information will constrain the search more specifically. In this method, additional parameters can be entered as parameters for the search, such as the mass of the parent ion (the peptide), as well as the enzyme used to generate this peptide from the original protein sample. Some examples of the computer search algorithms developed for this purpose include the MS/MS ion search tool, MASCOT (www.matrixscience.com), MS-TAG, which is available from the UCSF Mass Spectrometry Facility website at http://prospector.ucsf.edu/ucsfhtml4.0u/mstagfd.htm, TagIdent, available at http://ca.expasy.org/tools/tagident.html, and the SEQUEST search program, developed by J. Yates and J. Eng at the University of Washington. Positive protein identification can be made for an unknown if the sequence information in the CID spectra can be correlated to protein sequences available in the database. Also, if multiple peptides are present with CID spectra that correlate with the same protein, then a higher level of confidence in protein identification can be obtained. As with peptide mass fingerprinting, the scoring of the search results are based on similar forms of statistical treatment of the matching peptide fragments [61,62].

In the final case, the amino acid sequence of the peptide is unknown, but can be deduced directly from the fragmentation pattern. This is termed *de novo* sequencing, and is the most difficult case, as it requires that sufficient fragmentation information be available from the CID spectrum to fully determine the amino acid sequence of the peptide. However, this may be the only possible option, as is the case if the protein to be determined is not available in the protein sequence database.

1.3.2.3 Post-Translational Modification Analysis by MS. Once the amino acid sequence of a protein (its primary sequence) is determined, the next logical step in protein analysis is in the determination of post-translational modifications (PTM's), if they exist. More than 200 different forms of protein modifications have been reported [66]; common modifications include methylation, acetylation, phosphorylation, disulfide bond linkage, and glycosylation. At first glance, it would appear that MS based methods are ideally suited to determine the degree and type of post-translational modification present in a protein, since PTM's will alter the mass of the protein or peptide. However, determination of PTM's in proteins has proven to be notoriously difficult, as the natural abundance of modified proteins found in biological sources is often extremely low. As a result, the modified protein to be analyzed will be present in low quantity amidst a high background of the equivalent unmodified protein. To this end, purification or enrichment of the modified protein has been pursued to increase the relative abundance of modified protein (or peptides) in the sample.

PTM determination can be accomplished in several different ways [66]. One method involves CID or PSD fragmentation of peptides containing the modification. Characteristic fragments of the modifying residue can be used as markers to determine the exact site of modification. Another approach involves the use of enzymes that selectively cleave the modifying group (such as a phosphate group or a carbohydrate unit) from the peptide backbone. Mass analysis of the peptide before, and after enzyme processing can localize the nature and position of the modification.

1.4 Problems and Considerations in Sample Preparation

Given the power of MS-based methods for protein identification and characterization, mass spectrometry is now in widespread use for rapid protein identification on a high-throughput platform. Examples are available in the literature on the high-throughput identification of protein from various organisms [67-70]. However, the identification and characterization of unknown proteins is yet to be considered a routine procedure. One of the main factors which limits the ultimate success rate of protein identification is related to the preparation and manipulation of samples prior to mass analysis or fragmentation with MS instruments. Specifically, techniques related to the purification of contaminated samples, methods for protein preconcentration or enrichment, and issues related to the enzymatic degradation of samples (as required for peptide mass fingerprinting, for example) need to be optimized. These issues are discussed below.

1.4.1 Contaminants and Protein Mixtures. MALDI-TOF is characterized by its ability to analyze protein mixtures without separation [71]. In addition, the technique is tolerant to a certain degree of contaminants such as non-volatile salts [72]. The combination of these two aspects makes MALDI-MS ideally suited to the direct analysis of biological samples. As an example, MALDI has been extensively used to examine cell extracts from bacterial cultures, and this topic has been recently reviewed [73].

In all cases, protein samples obtained from biological sources require a certain amount of preparation before they can be analyzed. The extraction procedure often involves the addition of buffers, non-volatile salts, or surfactants to the sample. In addition, the raw sample may contain various salts and other components (lipids, carbohydrates, and nuclear material) along with the protein(s) of interest. When samples containing these components are analyzed by MALDI-MS, the contaminants can contribute to a loss in signal intensity, or in the worst case, complete signal suppression. Thus, many 'real world samples' will require further purification strategies prior to MS analysis, in spite of the high tolerance to contaminants associated with the MALDI technique [74].

1.4.2 Protein Concentration and Sensitivity Issues. Another concern with the preparation of protein samples is related to the protein concentration. MALDI-MS has proven to be extraordinarily sensitive technique for the analysis of peptides present in a pure system; an example has been reported on the detection of 25,000 molecules of Substance P from low volume deposition of samples by MALDI-MS [75]. However, although MALDI sensitivity is typically reported in terms of the molar quantity of sample analyzed, the upper limit of sample volume that can be analyzed is at the microliter level. Therefore, it is the concentration of protein present in the sample that will limit the ultimate sensitivity of MALDI analysis.

Although MALDI-MS is extremely sensitive for protein and peptide analysis, the sensitivity is affected by various factors. As mentioned previously, the presence of contaminating components in the sample can cause signal suppression. Also, intrinsic protein properties, such as the number of charges on the protein, the hydrophobicity, and even the molecular weight, will have a significant effect on the ionization efficiency in the MALDI technique [76].

Another concern with the protein concentration deals with the loss of protein during the various sample manipulation steps. With each step it is expected that there will be a certain amount of sample loss due to a less than 100% recovery. This can bring the level of protein below the minimal required concentration for analysis. Also, sample loss due to adsorption to containers becomes an increasing concern at low levels of sample present in solution.

1.4.3 Protein Degradation via Enzymatic Digestion. The concentration of protein in the sample will also limit enzyme digestion kinetics. Most MS-based strategies for protein identification, including peptide mass mapping and amino acid sequencing, first involve the generation of peptide fragments of the protein by sequence specific enzyme degradation. As with all reactions, the rate of enzyme proteolysis is dependant both on the amount of enzyme added to the system, and the concentration of protein in the sample [77].

To a first approximation, the reaction rate, ν, for a simple 1:1 enzyme-substrate system can be expressed as follows:

$$\nu = \frac{k_{cat}[E]_0[S]}{K+[S]}, \qquad (6)$$

where $[E]_0$ is the total concentration of enzyme, [S] is the concentration of substrate (the protein), K is the equilibrium constant of the enzyme-substrate complex, and k_{cat} is the turnover rate constant for the formation of product. Thus, by increasing the concentration of enzyme $[E]_0$, the rate of reaction can be increased. However, the presence of a high relative concentration of enzyme in the system as compared to the concentration of protein has a secondary effect. Since the enzyme is, in itself, a protein, the enzyme can act as its own substrate, thereby resulting in the production of autolytic peptide fragments. As the relative concentration of enzyme in the system increases, so to does the amount of enzyme autolysis. These autolysis fragments, if present in sufficient quantity, can partially or completely suppress the signals for the analyte(s) in a MALDI spectrum. Therefore, when the concentration of protein is extremely low, the enzyme concentration should also be kept reasonably low, and these low concentrations contribute to a slow reaction rate. Thus, there exists a concentration limit below which enzymatic degradation of the protein sample becomes difficult.

1.4.4 Methods to Improve MALDI Signals in Protein Analysis. Various strategies have been developed involving the manipulation of protein samples prior to MS analysis. These techniques attempt to reduce contaminants and impurities, increase protein concentration, or improve protein separation. In addition, several strategies are devoted to the digestion process involved with peptide mass fingerprinting.

The simplest strategy for sample cleanup can be directly related to the MALDI sample deposition method [26,32,35]. During crystallization of the matrix, water-soluble salts tend to precipitate last from the evaporating solvent. Thus an on-probe washing step can be performed to reduce the salt concentration by placing a droplet of water on top of the matrix spot. After a few seconds, most of the salts will have redissolved, and the water is removed. Other strategies for removing sample impurities such as non-volatile salts often involve the use of small cartridges containing ion exchange, or reversed-phase resins for selective removal of contaminants with minimal loss of sample. Millipore's commercial Ziptip cartridges are based on this principle, and are available in a variety of packing material depending on the user applications. Another interesting approach for sample cleanup involves modifying the MALDI target directly to allow for capture of protein samples with the ability to wash away weakly bound contaminants [77].

As MALDI-MS sensitivity is related to the concentration of protein in the sample, considerable effort is devoted to increasing the protein concentration prior to MS analysis. The simplest strategy involves evaporation of the solvent to increase protein concentration. However, this technique will also concentrate the contaminants, and there is also a risk of sample loss due to precipitation, or adsorption to the container walls. Other strategies involve the use of commercial cartridges with molecular weight cutoff filters to selectively concentrate protein samples while removing the impurities. Also, the same cartridges used in sample cleanup lend themselves to selective capture and concentration of proteins.

Enzymatic degradation remains a fundamental aspect of MS-based protein characterization. The choice of enzyme, or multiple enzymes is often key to the success of this method. Several novel procedures have been developed for protein degradation. These include the use of cartridges containing immobilized enzymes [78], strategies for digestion directly on-probe [79] and digestion in minute sample volumes [80].

Currently, the development of techniques associated with the preparation of samples for MS analysis, including separation, purification, preconcentration, and digestion remains an area of considerable research. The work presented here describes further advances in sample preparation strategies. In particular, this work presents methods for sample cleanup and preconcentration. Alternate approaches to protein digestion for peptide mass mapping will also be shown. In addition, this work addresses sample manipulation strategies for the enrichment of phosphorylated proteins.

1.5 Literature Cited

(1) Zubritsky, E. *Anal. Chem.* **2002**, *74*, 23A-26A.

(2) Mann, M.; Hendrickson, R. C.; Pandey, A. *Ann. Rev. Biochem.* **2001**, *70*, 437-473.

(3) McLuckey, S. A.; Wells, M. *J. Chem. Rev.* **2001**, *101*, 5711-606.

(4) Wilkins, M. R.; Sanchez, J. C.; Gooley, A. A.; Appel, R. D.; Humphery-Smith, I.; Hochstrasser, D. F.; Willians, K. L. *Biotechnol. and Genetic Eng. Reviews* **1995**, *13*, 19-50.

(5) Kellner, R. *Fresenius J. Anal. Chem.* **2000**, *366*, 517-524.

(6) Godovac-Zimmermann, J.; Brown, L. R. *Mass Spectrom. Rev.* **2001**, *20*, 1-57.

(7) Venter, J. C; *et al. Science* **2001**, *291*, 1304-1351.

(8) International Human Genome Sequencing Consortium *Nature* **2001**, *409*, 861-921.

(9) Peng, J.; Gygi, S. P. *J. Mass Spectrom.* **2001**, *36*, 1083-1091.

(10) Quadroni, M.; James, P. *Electrophoresis* **1999**, *20*, 664-677.

(11) O'Farrell, P. H. *J. Biol. Chem.* **1975**, *250*, 4007.

(12) Aebersold, R.; Goodlett, D. *Chem. Rev.* **2001**, *101*, 269-295.

(13) Aebersold, R. H.; Leavitt, J.; Saavedra, R. A.; Hood, L. E.; Kent, S. B. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6970-6974.

(14) Gevaert, K.; Vandekerckhove, J. *Electrophoresis* **2000**, *21*, 1145-1154.

(15) Vestal, M. L. *Chem. Rev.* **2001**, *101*, 361-375.

(16) Tanaka, K.; Ido, Y.; Akita, S. in *Proceedings of the Second Japan-China Joint Symposium on Mass Spectrometry*; Matsuda, H., Liang, X.-T., Eds.; Bando Press: Osaka, Japan, **1987**, 185-188.

(17) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151-153.

(18) Karas, M.; Bachmann, D.; Bahr, Y.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Processes* **1987**, *78*, 53-68.

(19) Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299-2301.

(20) Yalcin, T.; Wallace, W. E.; Guttman, C. M.; Li, L. *Anal. Chem.* (submitted, 2002).

(21) Karas, M.; Bachmann, D.; Hillenkamp, F. *Anal. Chem.* **1985**, *57*, 2935-2939

(22) Hillenkamp, F.; Karas, M.; Holtkamp, D.; Klüsener, P. *Int. J. Mass. Spectrom. Ion Process.* **1986**, *69*, 265-276.

(23) Hillenkamp, F.; Karas, M. *Int. J. Mass Spectrom.* **2000**, *200*, 71-77.

(24) Schreimer, D. C.; Li, L. *Anal. Chem.* **1996**, *68*, 2721-2725.

(25) Weinberger, S. R.; Boernsen, K. O.; Finchy, J. W.; Robertson, V.; Musselman, B. D. in *Proceedings 41st ASMS Conference on Mass Spectrom. Allied Topics;* San Francisco, CA, May 31-June 4, **1993,** pp 775a-b.

(26) Xiang, F.; Beavis, R. C. *Rapid Commun. Mass Spectrom.* **1994,** *8*, 199-204.

(27) Xiang, F.; Beavis, R. C. *Org. Mass Spectrom.* **1993,** *28*, 1424-1429.

(28) Mock, K. K.; Sutton, C. W.; Cottrell, J. S. *Rapid Commun. Mass Spectrom.* **1992,** *6*, 233-238.

(29) Bai, J.; Liu, Y. H.; Cain, T. C.; Lubman, D. M. *Anal. Chem.* **1994,** *66*, 3423-3430.

(30) Köchling, H. J.; Biemann, K. in *Proceedings 43rd ASMS Conference on Mass Spectrom. Allied Topics;* Atlanta, Georgia, May 21-26, **1995,** p. 1225.

(31) Hensel, R. R.; King, R.; Owens, K. G. in *Proceedings 43rd ASMS Conference on Mass Spectrom. Allied Topics;* Atlanta, Georgia, May 21- 26, **1995,** p. 947.

(32) Vorm, O.; Roepstorff, P.; Mann, M. *Anal. Chem.* **1994,** *66*, 3281-3287.

(33) Li, L.; Golding, R. E.; Whittal, R. M. *J. Am. Chem. Soc.* **1996,** *118*, 11662-11663.

(34) Kussmann, M.; Nordhoff, E.; Rahbek-Nielsen, H.; Haebel, S.; Rossel-Larsen, M.; Jakobsen, L.; Gobom, J.; Mirgorodskays, E.; Kroll-Kristensen, A.; Palm, L.; Roepstorff, P. *J. Mass Spectrom.* **1997,** *32*, 593-601.

(35) Dai, Y. Q.; Whittal, R. M.; Li, L. *Anal. Chem.* **1996,** *68*, 2721-2725.

(36) Karas, M.; Bahr, U.; Gieβmann, U. *Mass Spectrom. Rev.* **1991,** *10*, 335-357

(37) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1991,** *63*, 1193A-1203A.

(38) Beavis, R. C.; Chait, B. T. *Chem. Phys. Letters* **1993,** *181*, 470-484.

(39) Karas, M.; Glückmann, M.; Schäfer, J. *J. Mass Spectrom.* **2000,** *35*, 1-12.

(40) Kinsel, G.; Edmondson, R. D.; Russel, D. *J. Mass. Spectrom.* **1997,** *32*, 714-722.

(41) Karas, M.; Hillenkamp, F. *Organic Mass Spectrom.* **1992,** *27*, 472-480.

(42) Zenobi, R.; Knochenmuss, R. *Mass Spectrom. Rev.* **1998,** *17*, 337-366.

(43) Keller, B. O.; Li, L. *J. Am. Soc. Mass. Spectrom.* **2000,** *11*, 88-93.

(44) Cotter, R. J. *Anal. Chem.* **1992,** *64*, 1027A-1039A.

(45) Weickhardt, C.; Moritz, F.; Grotmeyer, J. *Mass Spect. Rev.* **1996,** *15*, 139-162.

(46) Stephens, W. E.; *Phys. Rev.* **1946,** *69*, 691.

(47) Wurz, P.; Gubler, L. Rev. Sci. Instrum. **1996,** *67*, 1790-1793.

(48) Cotter, R. J. in *Time-of-Flight Mass Spectrometry: Instrumentation and Applications in Biological Research*; A.C.S.: Washington, DC, **1997**.

(49) Whittal, R.; Li, L. *American Laboratory* **1997**, *28*, 30-36.

(50) Wiley, W. C.; McLaren, I. *Rev. Sci. Instrum.* **1955**, *26*, 115-1157.

(51) Whittal, R.; Russon, L.; Weinberger, S. R.; Li, L. *Anal. Chem.* **1997**, *69*, 2147-2153.

(52) Mamyrin, B. A.; Karatajev, V. J.; Shmikk, D. V.; Zagulin, V. A. *Soviet Phys. JETP* **1973**, *37*, 45-48.

(53) Medzihradsky, K. F.; Campbell, J. M.; Baldwin, M. A.; Falick, A. M.; Juhasz, P.; Vestal, M. L.; Burlingame, A. L. *Anal. Chem.* **2000**, *72*, 552-558.

(54) Loboda, A. V.; Krutchinsky, A. N.; Bromirski, M., Ens, W.; Standing, K. G. *Rapid Comm. Mass. Spectrom.* **2000**, *14*, 1047-1057.

(55) Van Berkel, W. J. H.; Van Den Heuvel, R. H. H.; Versluis, C.; Heck, A. J. R. *Protein Science* **2000**, *9*, 435-439.

(56) Nelson, R. W.; Dogruel, D., Williams, P. *Rapid Comm. Mass Spectrom,* **1995**, *9*, 625.

(57) Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5011-5015.

(58) Mann, M.; Hojrup, P.; Roepstorff, P. *Biol. Mass Spectrom.* **1993**, *22*, 338-345.

(59) Yates III, J. R.; Speicher, S.; Griffin, P. R.; Hunkapiller, T. *Anal. Biochem.* **1993**, *214*, 397-408.

(60) James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. *Biochem. Biophys. Res. Comm.* **1993**, *195*, 58-64.

(61) Pappin, D. J. C.; Hojrup, P.; Bleasby, A. J. *Curr. Biol.* **1993**, *3*, 327-332.

(62) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3351-3367.

(63) Mann, M.; Wilm, M. S. *Anal. Chem.* **1994**, *66*, 4390-4399.

(64) Spengler, B.; Kirsh, D.; Kaufmann, R.; Jaeger, E. *Rapid Commun Mass Spectrom.* **1992**, *6*, 105-108.

(65) Biemann, K.; Scoble, H. A. *Science*, **1987**, *237*, 992-998.

(66) Larsen, M. R.; Roepstorff, P. *Fresenius J. Anal. Chem.* **2000**, *366*, 677-690.

(67) Link, A. J.; Hays, L. G.; Carmack, E. B.; Yates III, J. R. *Electrophoresis*, **1997**, *18*, 1314-1334.

(68) Shevchenko, A.; Jenson, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Boucherie, H. Mann, M. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 1440-1445.

(69) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E., Mize, G. J.; Morris, D. R. Garvik, B. M, Yates III J. R. *Nat. Biotechnol.* **1999**, *17*, 676-682.

(70) Gygi, S. P., Rist, B., Gerber, S. A.; Turecek, F. Gelb, M. H., Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994-999.

(71) Beavis, R. C.; Chait, B. T. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 6873-6877.

(72) Mock, K. K.; Sutton, C. W.; Keane, A.; Cottrell, J. S. in *Techniques in Protein Chemistry IV.* Angeletti, R. H. Ed.; Academic Press, Inc. San Diego, CA; **1993**.

(73) Dalluge, J. J. *Fresenius J. Anal. Chem.* **2000**, *366*, 701-711.

(74) Yates III, J. R. *J. Mass Spectrom.* **1998**, *33*, 1-19.

(75) Keller, B. O.; Li, L. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1055-1063.

(76) Dai, Y.; Whittal, R. M.; Li, L. *Anal. Chem.* **1999**, *71*, 1087-1091.

(77) Price, N. C; Stevens, L. in *Fundamentals of Enzymology, The Cell and Molecular Biology of Catalytic Proteins, 3rdEd.* Oxford University Press, Oxford, **1999**.

(78) Hsieh, Y. L.; Wang, H.; Elicone, C.; Mark, J.; Martin, S. A.; Regnier, F. *Anal. Chem.* **1996**, *68*, 455-462.

(79) Owens, D. R.; Bothner, B.; Phung, Q.; Harris, K.; Siuzdak, G. *Bioorganic Med. Chem.* **1998**, *6*, 1547-1554.

(80) Whittal, R. M.; Keller, B. O.; Li, L. *Anal. Chem.* **1998**, *70*, 5344-5347.

Chapter 2

Two-Layer Sample Preparation Method for MALDI-Mass Spectrometric Analysis of Protein and Peptide Samples Containing Sodium Dodecyl Sulfate [1, 2]

2.1 Introduction

Matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) mass spectrometry (MS) are widely used for the analysis of biomolecules such as proteins and peptides. MALDI-MS is particularly attractive for the direct analysis of complex biological samples such as cell extracts [1]. This ionization technique can tolerate the relatively high levels of contaminating compounds, such as non-volatile salts, that are often present at significant levels in biological samples [2]. Such contaminants are often difficult to remove, either completely or partially, prior to MS analysis without extensive effort devoted to sample cleanup or separation. Detergents are a class of compounds that are often added to a sample during workup, with sodium dodecyl sulfate (SDS) being most commonly employed. SDS is needed in the commonly employed separation technique of polyacrylamide gel electrophoresis (SDS-PAGE). It is also used for improving protein solubilization, particularly for hydrophobic proteins, as well as reducing protein aggregation and minimizing sample loss due to adsorption to sample containers [3,4].

---

[1] A portion of this chapter has been published as: N. Zhang, A. Doucette, and L. Li "Two-Layer Sample Preparation Method for MALDI Mass Spectrometric Analysis of Protein and Peptide Samples Containing Sodium Dodecyl Sulfate" *Anal. Chem.* **2001**, *73*, 2968-2975.

[2] The bulk of the results described in this chapter were conducted with equal contributions from N. Zhang and A. Doucette. Specifically, for all data related to the investigation of the effects of SDS on MALDI, the results were recorded by both N. Zhang and A. Doucette, and subsequently averaged in order to provide an independent investigation of these effects.

Several reports have shown that SDS is detrimental to the MALDI signal of peptides and proteins and should therefore be removed from the sample prior to MS analysis, or avoided altogether [5-10]. Explored as a potential alternative to SDS, nonionic detergents such as β-D-octylglucoside are known to cause less signal degradation than SDS [6-8]. On the other hand, a few recent studies indicated that SDS could help the MALDI analysis. For example, Amado *et al.* showed that if the SDS concentration was increased above 0.23%, the ion signal was seen to improve [11]. They reported this phenomenon to be correlated with protein molecular weight, as well as with protein concentration. A partial signal recovery was consistently observed at higher SDS levels, and the concentration of SDS at which signal recovery occurred was very reproducible. Limbach *et al.* showed that the ion signal of hydrophobic peptides improved in the presence of SDS at levels above 0.5% [12]. While it is known that the addition of a detergent to the sample can reduce mass discrimination of peptide mixtures [10], SDS was recently shown to improve sequence coverage of proteins from enzymatic digests by improving the relative ionization of the more hydrophobic peptides [13,14].

We believe the wide range of results on the effects of SDS on MALDI-MS analysis might be explained by differences in the sample preparation methods used by various groups. It has been shown that pH and on-probe washing are crucial to the success of MALDI analysis of samples containing SDS, and that good spectra could be obtained for solutions containing up to 0.2% SDS [15]. However, it is not known if any other controllable variables for sample preparation are important when dealing with samples containing SDS.

One of the most influential variables is the technique of sample/matrix deposition on the MALDI sample plate. There are a number of sample preparation methods developed for MALDI applications. They include dried-droplet [16], vacuum drying [17], crushed-crystal [18], slow crystal growing [19], active film [20,21], pneumatic spray [22], electrospray [23], fast solvent evaporation [24,25], sandwich [26,27], and two-layer method [28]. Li's research group has recently demonstrated that the two-layer method can provide much improved performance in analyzing complex protein and peptide mixtures [29-30], compared to the conventional dried-droplet method and the fast evaporation method. The two-layer method involves the use of fast solvent evaporation to form the first layer of small matrix crystals, followed by deposition of a *mixture* of matrix and analyte solution on top of the crystal layer. With this method, the matrix and analyte solution conditions for preparing the second layer can be readily altered and fine-tuned for specific applications [29].

In this work, it is demonstrated that the two-layer sample preparation method is remarkably robust in handling protein and peptide samples containing SDS for MALDI analysis. A detailed study of the effects of SDS, under different preparation conditions involving the use of the two-layer method as well as the dried-droplet method, is performed. In addition, we present three new applications where the inclusion of SDS in the sample workup may be beneficial and where successful MALDI analysis was still possible in the presence of SDS using the two-layer method.

2.2 Experimental

2.2.1 Chemicals and Reagents. Bovine serum albumin (BSA), equine cytochrome c, myoglobin, ubiquitin, insulin chain B oxidized, bradykinin, and trypsin (TPKC treated to reduce chymotrypsin activity) were from Sigma Aldrich Canada (Oakville, ON). Sinapinic acid (SA), 2,5-dihydroxybenzoic acid (DHB), 2-(4-hydroxyphenylazo)-benzoic acid (HABA), α-cyano-4-hydroxycinnamic acid (HCCA), and sodium dodecyl sulfate were received from Aldrich Canada. HCCA was purified prior to use by recrystallization from ethanol. Analytical grade acetone, methanol, acetonitrile, formic acid, isopropanol, and trifluoroacetic acid (TFA) were purchased from Caledon Laboratories (Edmonton, AB). Water used in all experiments was from a NANOpure water system (Barnstead/Thermolyne).

2.2.2 MALDI Sample Preparation. Protein samples were prepared in SDS solutions and were thoroughly vortexed. The SDS concentrations (w/v) indicated throughout this work are for the original protein samples, *i.e.* prior to mixing with the MALDI matrices.

Two MALDI sample deposition methods were employed in this study, namely the dried-droplet, and two-layer method. For the dried-droplet method, the samples were mixed in a 1:1 (v/v) ratio with one of the four matrices studied, namely HCCA, HABA, SA, or DHB, and 1 μL of this mixture was deposited on a stainless steel target. With HCCA, HABA, and sinapinic acid, the matrix was saturated in 50% (v/v) acetonitrile/water unless otherwise indicated. With DHB, the matrix was dissolved in 9:1 (v/v) water/ethanol at a concentration of 20 g/L.

For the two-layer method, 1 μL of a 12-mg/mL solution of HCCA in 80% (v/v) acetone/methanol was deposited on the target and was allowed to rapidly dry, thus forming a very thin layer of fine crystals (subsequently referred to as the first layer). Next, a 1 μL aliquot of the protein solution was mixed with the appropriate volume of HCCA solution (saturated in 40% methanol/water) and a 0.4 μL portion of this mixture was deposited on top of the first layer and allowed to dry.

For each method, an on-probe washing step was performed in cases where the sample spot (*i.e.* the analyte/matrix crystal) was not water-soluble. The washing step involves adding 1 μL of water to the sample spot and then absorbing the water with a Kimwipe. This procedure was performed twice for each sample spot.

2.2.3 Bacteria Extraction. The *Escherichia coli* used in this study (*E. coli* ATCC 9637) was cultured at the Edgewood RDE Center (EDREC). A 1 mg sample of *E. coli* was mixed with 500 μL of water. The sample was briefly vortexed to suspend the cells, and the suspension was divided evenly among 5 separate vials. To remove excess salts, the suspensions were vortexed for a few seconds, and the supernatant was discarded after centrifugation. The *E. coli* samples were extracted with 0.1% TFA by vortexing for 3 minutes followed by centrifugation. The pellet was then further extracted with 0.1% TFA containing various concentrations of SDS. The supernatant of the extracts was collected for MALDI analysis after mixing 1:1 with saturated HCCA solution (in a 1:2:3 volume ratio of formic acid, isopropanol and water). Samples were stored on ice prior to MALDI analysis.

2.2.4 Enzyme Digests. Protein samples (1 μg/μL) were prepared in 0%, 0.1% or 1% SDS solutions. Fourteen-microliter aliquots of the protein solutions were mixed with 1.5 μL of 1.0 M NH_4HCO_3 (to give a pH of ~8) and 1.4 μL of 1 μg/μL trypsin in siliconized vials. The samples were incubated at 37°C for 2 h. Following incubation, a 1-μL portion of the digested sample was mixed with 9 μL of saturated HCCA solution in 40% (v/v) methanol/water for MALDI detection. A control experiment was also performed by digesting protein samples prepared in water, then adding SDS to the digests to produce samples containing 0%, 0.1% or 1% SDS.

2.2.5 Instrumentation. MALDI experiments were carried out on a HP Model G2025A MALDI-TOF-MS (Hewlett-Packard, Reno, NV) or an Applied Biosystems Voyager Elite laser desorption/ionization time-of-flight (TOF) mass spectrometer (Framingham, MA). The linear mode of operation was used for protein detection while the reflectron mode was for peptide detection. Ionization was performed with a 337-nm pulsed nitrogen laser. The laser power was adjusted slightly above the threshold of the desorption/ionization process and was kept constant for a given set of protein samples (*i.e.*, one protein at a given concentration, with various concentrations of SDS). All spectra resulted from signal averaging of 50 to 64 shots. For each sample, the laser was moved throughout the sample spot in order to generate the best possible spectrum in each case.

MS/MS fragmentation spectra were recorded on an MDS Sciex API Q-Star quadrupole-time-of-flight mass spectrometer equipped with a MALDI source (Toronto, ON).

All data were processed using the Igor Pro Software package (WaveMetrics, Lake Oswego, OR). Signal to noise ratios (S/N) were calculated from the peak height (the signal) and the standard deviation of the baseline (noise), determined at a region near the peak of interest. Using the Igor Pro software, the peak heights were determined from the difference between the baseline, calculated as the average over the region used to calculate S/N, and the highest point of a Gaussian fitted curve of the peak. The error bars in the figures represent the calculated standard deviation of the signal to noise ratios of the spectra. At least three spectra, recorded from independent sample spots, were averaged for each data point in the figures.

2.3 Results and Discussion

2.3.1 Effect of On-Probe Sample Washing. The sample/matrix crystals prepared by the two-layer method can be readily washed with water. The effect of sample washing on MALDI analysis of proteins containing different amounts of SDS is illustrated in Figure 2.1. For this experiment, samples consisting of 500 nM BSA in 0%, 0.05%, or 1% SDS were prepared. The samples were mixed in a 1:4 ratio with saturated HCCA matrix solution and spotted on a stainless steel probe coated with the first layer of matrix crystals. The samples were initially analyzed by MALDI-MS without performing the on-probe washing step. Following acquisition of the spectra, the same sample spots were then washed on-target and re-analyzed by MALDI-MS. Figure 2.1 displays some representative MALDI spectra obtained from 500 nM BSA samples with and without on-probe washing.

As shown in Figure 2.1A, a strong $[M+H]^+$ BSA signal was observed for the sample prepared without SDS. The calculated signal to noise (S/N) ratio of this peak is 23. However, the presence of 0.05% SDS in the sample resulted in a poorly detected peak, with a calculated S/N of ~3 (see Figure 2.1B). At all levels of SDS higher than 0.1%, the BSA peak was undetectable, as shown in Figure 2.1C at an SDS concentration of 1%. When the same sample spots were washed on-probe, a remarkable difference was observed. Figure 2.1D shows that on-probe washing reduces the signal intensity for the sample that does not contain SDS. In this case, for a sample spot being essentially free of contaminants, a loss in signal likely results from a loss of sample occurring during the process of on-probe washing. In Figure 2.1E, for the BSA sample prepared in 0.05% SDS, again a weak BSA peak was observed, similar in intensity to the spectrum obtained prior to on-probe washing (Figure 2.1B). However, for the BSA sample prepared in 1% SDS, although no signal was obtained from the sample spot analyzed without a washing step (see Figure 2.1 C), a strong BSA signal, with an S/N ratio of 12, was detected from the same sample spot on-probe washing (see Figure 2.1F). These results were also obtained with various other protein samples, including cytochrome c, lactoferrin, and insulin chain B, using HCCA as matrix. The deterioration in signal at low levels of SDS and signal recovery at higher SDS levels is in agreement with earlier work presented by Amado *et al.* [11].

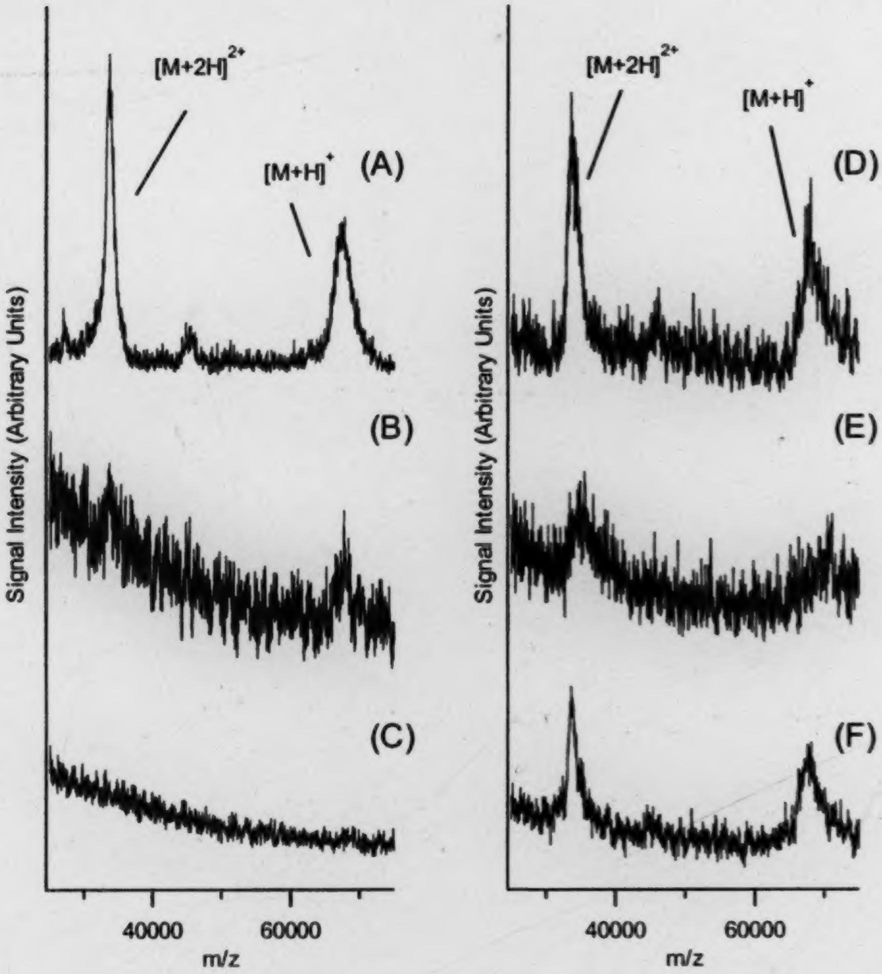

Figure 2.1 MALDI spectra obtained from samples of 500 nM BSA prepared in (A) 0% SDS (S/N=23), (B) 0.05% SDS (S/N=3), (C) 1% SDS (S/N=0), without on-probe washing; and in (D) 0% SDS (S/N=8), (E) 0.05% SDS (S/N=3), (F) 1% SDS (S/N=12), with the addition of on-probe washing. (A) & (D), (B) & (E), and (C) & (F) were from the same sample spot. The indicated signal to noise ratios are for the singly protonated ion of BSA: [M+H]$^{+}$.

In summary, the above results indicate that a washing step is crucial in order to obtain a signal at high SDS concentrations (>0.1%) with the two-layer sample preparation method. It should be noted that although the washing step appears to improve MALDI signals at high SDS concentrations, not all sample preparation conditions are tolerant to on-probe washing. It was observed that as the SDS level was increased beyond 1%, the sample spots became increasingly soluble during the washing step. In addition, the water-soluble matrix, DHB will completely dissolve if subjected to such washing. In general, however, a washing step should be performed on all samples if the deposited spot can tolerate this procedure.

2.3.2 Effect of Matrix. Four commonly used matrices, namely HCCA, DHB, HABA, and SA, were employed in this study. Solutions of 1 μM cytochrome c prepared with various concentrations of SDS were analyzed using these four matrices. We note the two-layer method is not applicable to DHB – this matrix does not dissolve in an appropriate organic solvent such as acetone, which is required to form a thin layer by fast evaporation. To include DHB in this comparative study, the dried droplet deposition method was used for all matrices. On-probe washing was performed on the samples, except for those prepared with DHB or with HABA at SDS levels ≥1%, as these sample spots were completely water soluble. Figure 2.2 summarizes the average signal to noise ratios of the singly charged cytochrome c peaks obtained from the MALDI spectra using each matrix at varying concentrations of SDS.

Figure 2.2 shows that the type of matrix used plays a significant role in determining spectral quality at varying concentrations of SDS. Based on these results, sinapinic acid was least tolerant towards the presence of SDS in the sample. Figure 2.2A shows that for SA as matrix, at most levels of SDS, only a very weak signal, or a

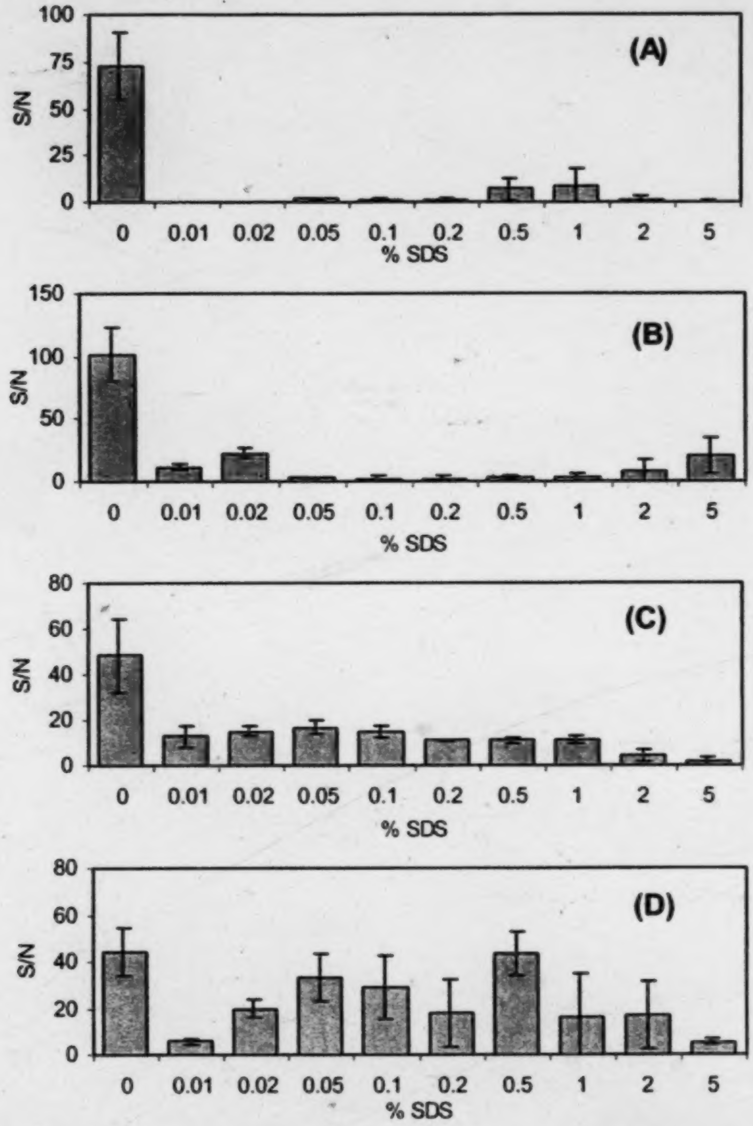

Figure 2.2 Average signal to noise ratios (S/N) for the molecular ion signals calculated from the MALDI spectra of 1 μM cytochrome c at various concentrations of SDS. The matrices used for the analysis are (A) sinapinic acid, (B) HCCA, (C) HABA, (D) DHB. Error bars represent the standard deviations of S/N from 9 spectra recorded from 3 independent sample spots (A, C, D) or 3 independent spectra (B).

complete lack of signal for cytochrome c was observed. The spectra acquired using HCCA as matrix also display generally weak signals when SDS was added to the samples (Figure 2.2B). The matrix HABA seemed to be more tolerant to SDS (Figure 2.2C), with good S/N at all levels of SDS ≤1%. A decrease in signal intensity was seen for HABA at high levels of SDS (2% and 5%). Finally, samples prepared using DHB as matrix yielded much stronger signals in the presence of SDS as compared to the other matrices (Figure 2.2D). However, DHB tends to form a characteristic ring of crystals at the edge of the sample spot, giving poor spot-to-spot spectral reproducibility. It was necessary to search for "hot" spots in order to obtain any detectable signal.

It was not surprising that the choice of matrix would play a role in obtaining MALDI spectra for SDS-containing samples. The success of protein ionization and subsequent detection depends extensively on co-crystallization of sample and matrix [28]. The addition of SDS to the sample is expected to change the solubility of the protein and likely affects the extent of analyte incorporation as well as analyte distribution in the matrix crystals. However, the results presented in Figure 2.2 cannot be exclusively correlated to the choice of matrix, since other factors are expected to influence the trends. In particular, the solvent system was not identical for all matrices used. Sinapinic acid, HABA, and HCCA were prepared as saturated solutions in 50% acetonitrile/water, while the DHB solution was prepared as a 20-mg/mL solution in 9/1 water/ethanol. Therefore, we investigated the effects of solvent and sample preparation method on SDS-containing samples using HCCA as matrix. HCCA was chosen since it provides superior detection sensitivity and higher tolerance to salts for protein samples, particularly high molecular weight proteins [33], compared to SA and HABA.

2.3.3 Effects of Solvent and Sample Preparation Method. Samples of 1 μM cytochrome c were prepared and analyzed by MALDI-MS using two different sample preparation methods with HCCA as the matrix. In the dried droplet method, the sample was mixed with HCCA (saturated in either 50% acetonitrile/water or 40% methanol/water) in a 1:1 ratio. For the two-layer method, the sample was mixed in a 1:1 ratio with HCCA (saturated in either 50% acetonitrile/water or 40% methanol/water) and 0.4 μL of the mixture was deposited on top of the first layer of matrix crystals. All samples were washed on-probe and then analyzed by MALDI-MS. Figure 2.3 displays the averaged signal to noise ratios calculated from the MALDI spectra collected under the various conditions.

The comparison of the results between Figure 2.3A and 2.3B indicates that, for the dried-droplet method, the solvent system has a significant effect on the protein signals at different levels of SDS. Figure 2.3A was obtained by using the sample prepared with HCCA in 50% acetonitrile/water and it shows that the analyte signals were reduced drastically for the samples containing different levels of SDS. In contrast, Figure 2.3B illustrates that when the sample was prepared in the methanol/water solvent system, at all levels of SDS up to 2%, the S/N ratio was equal to or greater than that of the sample prepared without SDS. The cause of this strong solvent effect is unknown. It might be related to the extent of analyte incorporation into matrix crystals prepared under different solvent conditions.

Although the choice of solvent has a significant effect on signal response for SDS-containing samples deposited using the dried droplet method, this effect was not observed when a two-layer sample preparation method was used. This can be seen from

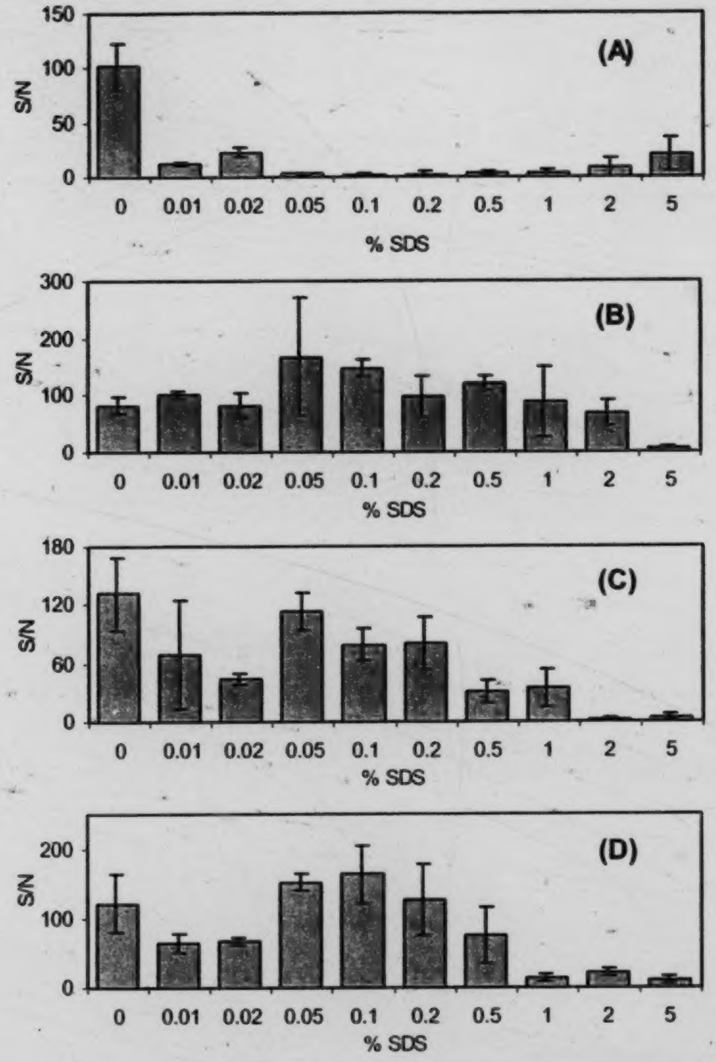

Figure 2.3 Average signal to noise ratios (S/N) for the molecular ion signals calculated from the MALDI spectra of 1 μM cytochrome c at various concentrations of SDS using HCCA as matrix with different sample deposition methods and solvent systems. The dried droplet method was used with HCCA saturated in (A) 50% acetonitrile/water and (B) 40% methanol/water. The two-layer method was used with the second layer HCCA saturated in (C) 50% acetonitrile/water and (D) 40% methanol/water. Error bars represent the standard deviations of S/N from 3 independent spectra.

the comparison of Figure 2.3C and 2.3D. In either of the two solvent systems, the addition of 0.01% or 0.02% SDS was seen to reduce the observed S/N ratio of the cytochrome c peak using a two-layer method. The presence of 0.05% SDS in the sample resulted in signal recovery with each solvent system. When the SDS concentration was further increased, the S/N ratio of the cytochrome c peak deteriorated.

The results of Figure 2.3 also show that, under identical solvent conditions, the sample deposition method has an effect on the signal intensity of SDS-containing samples. This is most clearly seen for the samples prepared using the acetonitrile-water solvent system, shown in panels A and C of Figure 2.3. Large differences are noted at levels of SDS in the range of 0.05% to 0.2%. At these levels of SDS, very weak signals were obtained with the dried droplet method (Figure 2.3A). However, using the two-layer method, the addition of 0.05 to 0.2% SDS resulted in signals with intensities that were comparable to the signal for the sample without SDS (Figure 2.3C). The samples prepared using the methanol-water solvent system, shown in Figure 2.3B and 2.3D, also show the dependence of signal intensity on sample deposition method. Significant differences are noted for the samples prepared in 1% and 2% SDS. While the dried droplet method still revealed relatively strong signals at these high levels of SDS, the samples prepared using the two-layer method displayed only weak signals as compared to the samples prepared without SDS.

It should be noted that, when the dried droplet method was used for the samples prepared at SDS levels above 1%, it became increasingly difficult to find the hot spots that produced a cytochrome c peak with high signal to noise ratios, as shown in Figures 2.3A and B. The ability to collect MALDI spectra with strong analyte signals for samples containing 2% or 5% SDS can be attributed to the non-homogeneous analyte distribution in the matrix/analyte crystals prepared by using the dried droplet method. Some regions of the sample spot, albeit difficult to find, contain the proper crystals that give rise to the cytochrome c signals. For general applications in analyzing samples containing SDS up to ~1%, the two-layer sample preparation is preferred. With this method, the analyte signals can be easily observed without the need of searching for the hot spots.

2.3.4 Effect of Protein Concentration. Another important variable in MALDI sample preparation is the protein concentration, which determines the analyte to matrix ratio in the co-crystals formed on the probe. For SDS-containing samples, protein concentration also governs the relative amount of SDS to protein for a given concentration of SDS. Figure 2.4 displays the average signal to noise ratios for the singly charged molecular ion signals obtained from the MALDI spectra of 500 nM, 1.5 μM, 5 μM, and 15 μM BSA with various concentrations of SDS. The sample spots were prepared using the two-layer method with the second layer containing analyte/matrix in a 1:4 ratio.

Figure 2.4 shows different trends for the low and high concentrations of BSA. For the 500-nM solution (Figure 2.4A), one can see that the MALDI signal intensity dropped only slightly at low amounts of SDS, and then fully recovered as the SDS concentration was increased to 0.5%. As the SDS concentration was increased above 0.5%, the signal intensity again dropped. Figure 2.4D shows that, for a 15 μM solution of BSA, the signal intensity decreased significantly as the concentration of SDS in the sample increased from 0% to 0.05%. The signal intensity did not return to any significant extent as the SDS concentration was further increased. At intermediate levels between these low and high concentrations of BSA, one can observe that for a 1.5 μM solution of BSA (Figure 2.4B), the trend most resembled that of the 500 nM BSA solution,

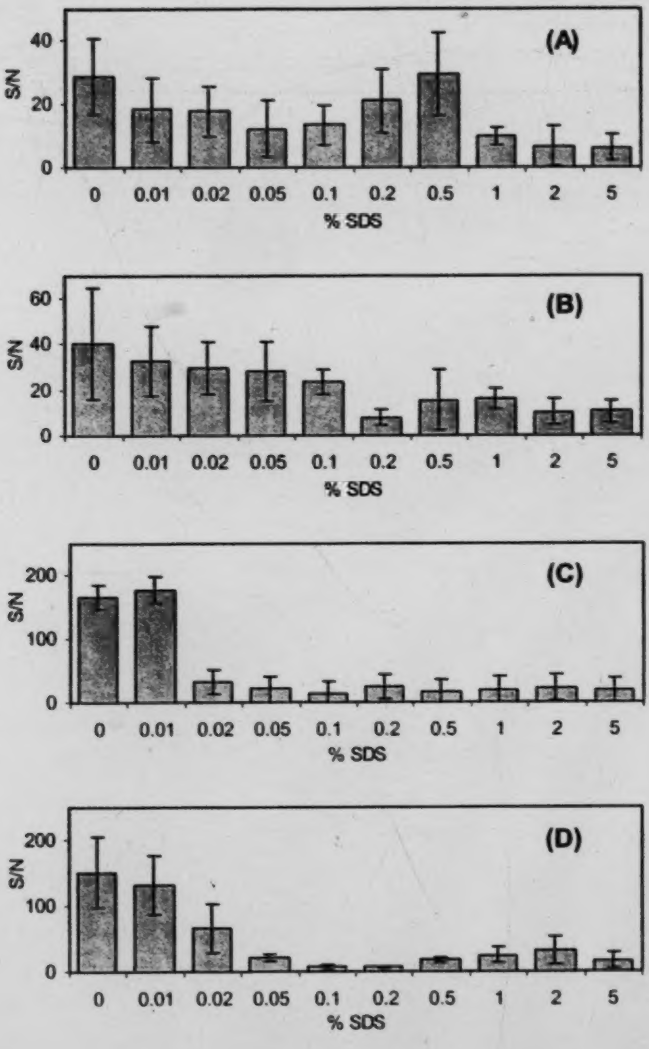

Figure 2.4 Average signal to noise ratios (S/N) for the molecular ion signals obtained from the MALDI spectra of BSA at various concentrations of SDS. (A) 500 nM, (B) 1.5 μM, (C) 5 μM, or (D) 15 μM BSA was mixed with the second-layer matrix solution in a ratio of 1:4, and (E) 15 μM BSA was mixed in a ratio of 1:120. The second-layer matrix solution consisted of 40% methanol/water HCCA matrix. Error bars represent the standard deviations of S/N from 3 independent spectra.

with only a slight drop in signal intensity as the SDS was increased to 0.1%. The 5 µM BSA samples (Figure 2.4C) displayed a trend that was very similar to the 15 µM BSA solution. These results demonstrate that protein concentration in the sample has an effect on the observed signal trends at different levels of SDS concentrations.

While Figure 2.4 clearly shows that the protein concentration will affect the signal intensities at different levels of SDS, the origin of this affect cannot be determined from this single experiment. The effect may result from variations in one, or a combination of several variables between each set. The most notable variations include the ratios between protein and matrix, as well as protein to SDS. These variations are independently explored in the following section.

2.3.5 Effect of Protein to Matrix Ratio. To address how the protein to matrix ratio in the presence of different amounts of SDS can affect the protein signal intensity, solutions of 15 µM BSA were prepared with various concentrations of SDS (0%, 0.5%, 1%, and 5%). The samples were mixed with matrix solutions at ratios of 1:1, 1:4, 1:10, 1:50, and 1:100. All samples were analyzed by MALDI-TOF MS using identical laser power. The average signal to noise ratios of the singly charged BSA peaks were calculated from the spectra and the results are summarized below in Table 2.1.

Table 2.1 Signal to noise ratios (S/N) for 15 μM BSA samples prepared with various concentrations of SDS at different protein to matrix ratios.

Protein to matrix ratio	Calculated S/N			
	0% SDS	0.5% SDS	1% SDS	5% SDS
1:1	29 ± 11	85 ± 20	68 ± 24	21 ± 4
1:4	215 ± 39	55 ± 28	55 ± 15	44 ± 5
1:10	240 ± 14	32 ± 10	35 ± 9	36 ± 16
1:50	107 ± 22	44 ± 14	37 ± 8	15 ± 3
1:100	27 ± 6	55 ± 15	39 ± 4	7 ± 5

From Table 2.1, the effect of varying the protein to matrix ratio on the signal intensity is clearly observed. Going down the column for the BSA sample prepared without SDS (0%), the strongest signals were observed for the samples that were diluted 1:4 or 1:10 with matrix. As might be expected, at low dilutions of sample to matrix (1:1), there is insufficient matrix on the target to effectively ionize the high quantity of BSA deposited on the probe (3 pmol). At high dilutions of sample to matrix (1:50, 1:100), the decreased amount of BSA present on the target results in a loss in signal intensity.

For the SDS-containing samples of Table 2.1, we can also observe a signal intensity dependence on protein to matrix ratio; however, this dependence is less pronounced than that of the sample without SDS. By comparing the S/N ratios at 0.5%, 1% or 5% SDS (down a given column of Table 2.1), only a small difference in S/N is incurred by changing the ratio of protein to matrix.

Another important observation can be made by comparing the S/N ratios obtained for a fixed ratio of protein to matrix (going across a row of Table 2.1). From Table 2.1, we can see that at the 1:1 dilution of protein to matrix, the signal to noise ratio was seen to increase from 29 at 0%, to 85 at 0.5% SDS; in other words, an improvement in signal intensity was achieved by adding SDS to the sample. When 1% SDS was added to the sample, the S/N ratio still remained high, and at 5% SDS, the signal returned to an intensity similar to the 1:1 dilution without SDS. A very similar trend was also obtained for the 1:100 dilution of sample with matrix, with signal improvement being achieved with the addition of 0.5% and 1% SDS to the sample. However, from the results in Table 2.1, a significantly different trend was obtained for the 1:4 and 1:10 dilutions of sample with matrix. From these dilutions, the strongest signal was observed for the sample containing no SDS. When SDS was added to the sample at the 1:4 and 1:10 dilutions, a marked decrease in signal intensity was observed. For a given dilution, the S/N ratio obtained at the 0.5%, 1%, and 5% levels of SDS were essentially identical. A similar trend is also apparent at the 1:50 dilution, where again the addition of 0.5% SDS or more to the sample significantly decreased the signal intensity. It is therefore apparent that the type of trend obtained by increasing the SDS in the sample will depend on the optimization of the protein to matrix ratio.

The difference in these two general trends can be related to the optimization of protein to matrix ratio. When the sample was mixed with matrix at or near the optimal ratio of 1:10, the addition of SDS to the sample results in decreased signal to noise. These results are in agreement with the theory that SDS decreases signal intensity, presumably because it interferes with the crystallization process [6]. However, for those samples that were prepared with a 1:1 or a 1:100 dilution of sample to matrix, adding SDS to the sample results in an improvement in signal. These S/N ratios, although greater than those for the sample without SDS at an equivalent analyte to matrix ratio, are still significantly lower than the optimal signal to noise ratio for the sample without SDS.

We note that the reason(s) why the presence and the amount of SDS in a protein and peptide sample can affect MALDI signals is unknown. SDS can affect sample preparation and/or ionization. In the two-layer method, the analyte/matrix crystals are in sub-micron sizes; there are no visible differences (under 50× magnification) in the morphology of crystals prepared with varying amounts of SDS. One can speculate that proteins are still associated with SDS in the matrix crystals; however the specific association of matrix, protein and SDS in the crystals is unknown. The conformation of the protein in the crystal is also unknown. SDS, present in solution at sufficient concentration, is well known to cause protein denaturation. However, the process of crystallization with the matrix may again change protein conformation. It is thus difficult to ascertain if protein denaturation by SDS plays any role on MALDI signal detectability.

Based on the above study, it is quite clear that the two-layer method, using well-controlled sample preparation conditions, can be used to handle protein samples containing very high concentrations of SDS. However, although signals were obtained for protein samples, the spectral quality of SDS-containing samples was affected. In particular, an intense background in the low molecular weight region was observed at high SDS concentrations. This is explored in the next section.

2.3.6 Blank Solutions and SDS Adduct Formation. Samples of SDS were prepared in water, and analyzed by MALDI-MS with different matrices. It was found that at concentrations of SDS exceeding 0.5%, strong clusters in the low-mass region of the spectrum became increasingly evident. This is in contrast to a previous report, which stated that even for a 10% solution of SDS, no interfering ions were detected in the low mass region until laser energies were well above threshold [12]. In our case, these interfering ions were observed at all laser powers that were sufficient to produce any matrix ion signal. Figure 2.5 shows the MALDI spectra obtained from "blank" solutions containing 1% SDS using HABA, DHB, and HCCA as matrix. Several peaks are observed in the mass region from 0 to 2500. It was observed that several of the detected peaks were spaced at mass intervals of 288.4 Da, which corresponds to the mass of SDS. Several other peaks were spaced at 22 Da, indicating sodium ion attachment. Based on peak mass alone, the signals could be attributed to clusters of SDS and sodium ions, along with attachments to matrix ions or matrix fragment ions.

The peak assignments of the low mass ions in the blank HCCA sample were confirmed by MALDI-MS/MS on a Q-TOF instrument. The results of these experiments can be summarized in Figure 2.6. In Figure 2.6A, a total ion scan was recorded from m/z 100 Da to 1500 Da. As already observed in Figure 2.5 using a Voyager Elite TOF instrument to record the spectra, several intense peaks of mass difference 288.4 Da were also recorded using the Q-TOF instrument. From the total ion scan, a parent ion of mass 1463 Da was chosen for fragmentation. This spectrum is displayed in Figure 2.6B. This MS/MS spectrum displays several intense fragments, whose masses are labeled in the figure. These fragments correspond to successive neutral losses of a fragment with mass 288 Da, which corresponds to the mass of SDS. The smallest fragment seen in this MS/MS spectrum, with mass 311 Da, corresponds to the mass of SDS with a single sodium ion attachment. These assignments were further confirmed by selecting a second cluster ion originally observed in the single MS scan. The MS/MS spectrum for the parent ion of mass 887 Da is displayed in Figure 2.6C. Again, several fragments were obtained corresponding to successive losses of 288 Da from the parent ion. Two additional fragments were also detected in this spectrum, with mass 142 Da, and 165 Da. The ion of mass 142 Da can be assigned to the radical cationic species $[SO_4Na_2]^+$, produced from homolitic cleavage of the SDS. The second ion of mass 165 Da

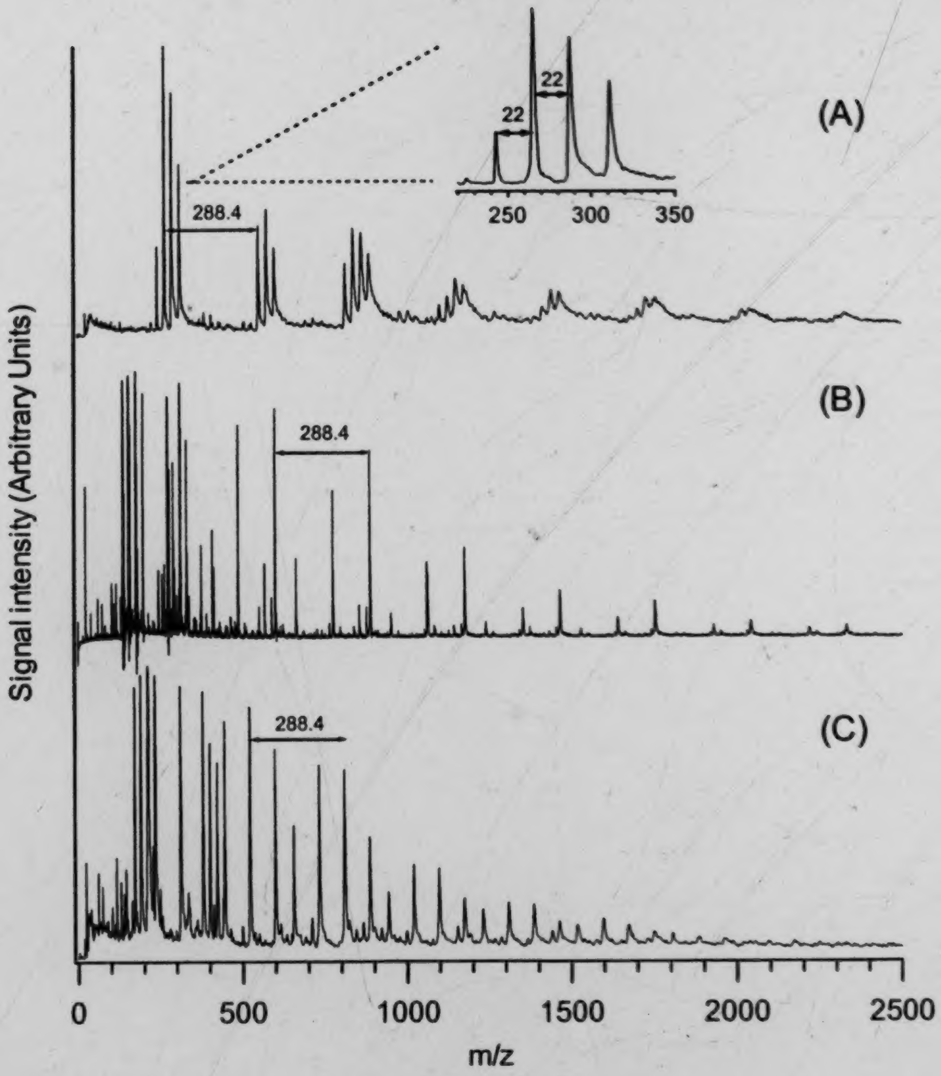

Figure 2.5 'Blank' MALDI spectra obtained from solutions of 1% SDS with (A) HABA, (B) DHB, and (C) HCCA as matrix. Representative peaks with mass differences of 288.4 Da, or 22 Da, are labeled in the figure.

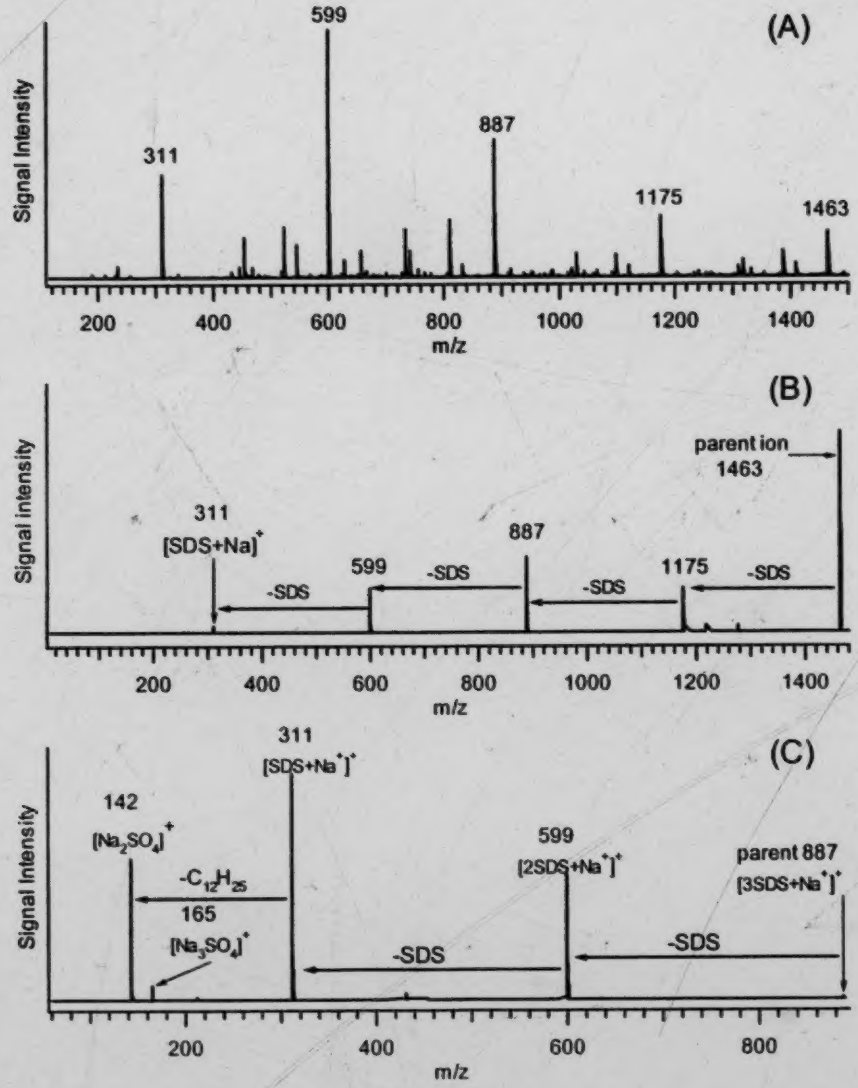

Figure 2.6 (A) Shown is a total MS scan of a 'blank' MALDI spectrum of a solution of 1% SDS with HCCA as matrix, recorded on a Q-star instrument. The MS/MS spectrum displayed in (B) was recorded by fragmentation of the ion from (A) with mass 1463 Da, and the MS/MS spectrum in (C) is from fragmentation of the ion with mass 887 Da.

corresponds to a sodium attachment to the neutral species $[Na_2SO_4+Na]^+$.

The SDS clusters shown in Figure 2.6 were also present in the low mass region of the protein samples, at SDS levels beyond 0.5% to 1%. At higher laser powers, the clusters resulted in a significant rise in the low mass background, and these SDS clusters were detected at masses up to 5000 Da. When the low mass peptide bradykinin was analyzed, interfering ions were found close to or overlapping the bradykinin peak, depending on the choice of matrix. These intense SDS cluster ions can make peak detection difficult, especially for peptide analysis.

Based on the effects of variables mentioned above, it would seem difficult to determine a standard set of conditions for MALDI sample preparation given that a sample contains a certain concentration of SDS. However, it was observed that the two-layer method could, in general, provide intense MALDI signals for protein samples, even at high concentrations of SDS. We therefore look at cases where the addition of SDS might be beneficial in sample workup prior to MS analysis. In the following sections, we demonstrate some applications of protein sample preparations involving SDS and study the effect of SDS on MALDI signals in these applications.

2.3.7 Hydrophobic Protein. Many biologically significant membrane-associated proteins are hydrophobic. Surfactants are often added to solutions of hydrophobic proteins in order to facilitate solubilization, as well as to avoid sample loss due to adsorption to the walls of the container. In this study, bacteriorhodopsin was used as a model hydrophobic protein. In Figure 2.7, the MALDI spectra obtained from the analysis of 0.25 μg/μL bacteriorhodopsin solutions containing various levels of SDS are shown. The two-layer method was employed, using HCCA as matrix. Figure 2.7A corresponds to the sample prepared in aqueous solution without SDS. It shows the bacteriorhodopsin

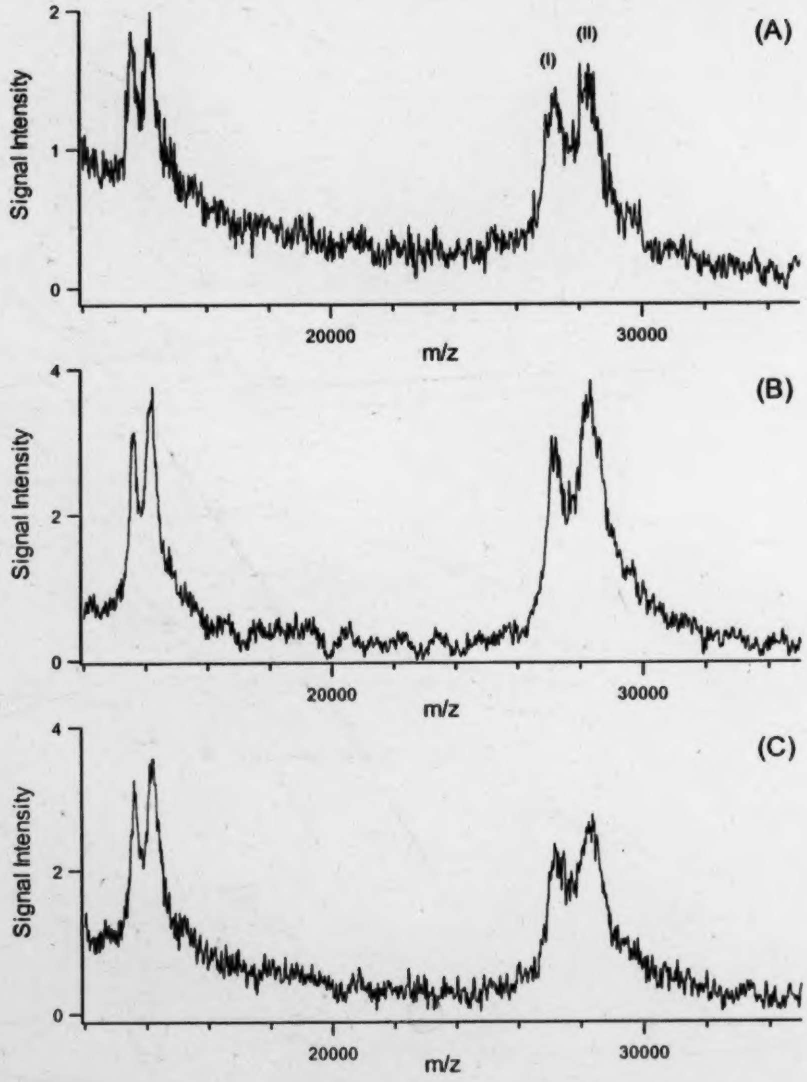

Figure 2.7 MALDI spectra obtained from samples of 0.25 μg/μL bacteriorhodopsin with (A) 0% SDS, (B) 0.1% SDS and (C) 1% SDS using HCCA matrix and a two-layer deposition method. The labeled peak I indicates the bacteriorhodopsin molecular ion signal and the peak labeled as II is likely from an unprocessed precursor of bacteriorhodopsin.

peak near m/z 27000, along with a second peak approximately 1200 Da higher than the bacteriorhodopsin peak, which likely corresponds to an unprocessed precursor of bacteriorhodopsin [34]. When 0.1% SDS was added to the sample, the resulting spectrum (Figure 2.7B) gave a stronger signal. At the 1% SDS level, the bacteriorhodopsin peak is still clearly detected. An increase in signal intensity obtained by adding SDS to the sample is perhaps due to the improved solubility of the protein with the SDS solvent system. This representative example demonstrates that hydrophobic proteins can be analyzed by the two-layer sample preparation MALDI-MS in the presence of SDS.

2.3.8 Bacterial Protein Extraction. Detergents such as SDS are often incorporated in extraction procedures to solubilize certain types of proteins from cells, such as the more hydrophobic membrane proteins [3]. Here we demonstrate that the addition of SDS to *E. coli* protein extraction aids in the detection of additional proteins by MALDI. In this case, a 2 mg/ml sample of *E. coli* was first extracted with 0.1% aqueous TFA in order to solubilize the more water-soluble proteins. Figure 2.8A depicts the MALDI spectra obtained from the analysis of this extract. Several peaks are visible in the spectrum and they are mainly from proteins [30,31]. The *E. coli* cell pellet was further extracted using an identical solvent. This extract was then analyzed by MALDI-MS and the spectrum is shown in Figure 2.8B. Very weak signals were obtained, indicating that most of the water-soluble proteins were already removed by the first extraction. The pellet was then further extracted using a solution containing 0.2% SDS in 0.1% aqueous TFA. Figure 2.8C shows the MALDI spectrum from this extract. A strong peak corresponding to a protein with MW~7000 Da is visible, along with the doubly charged molecular ion peak.

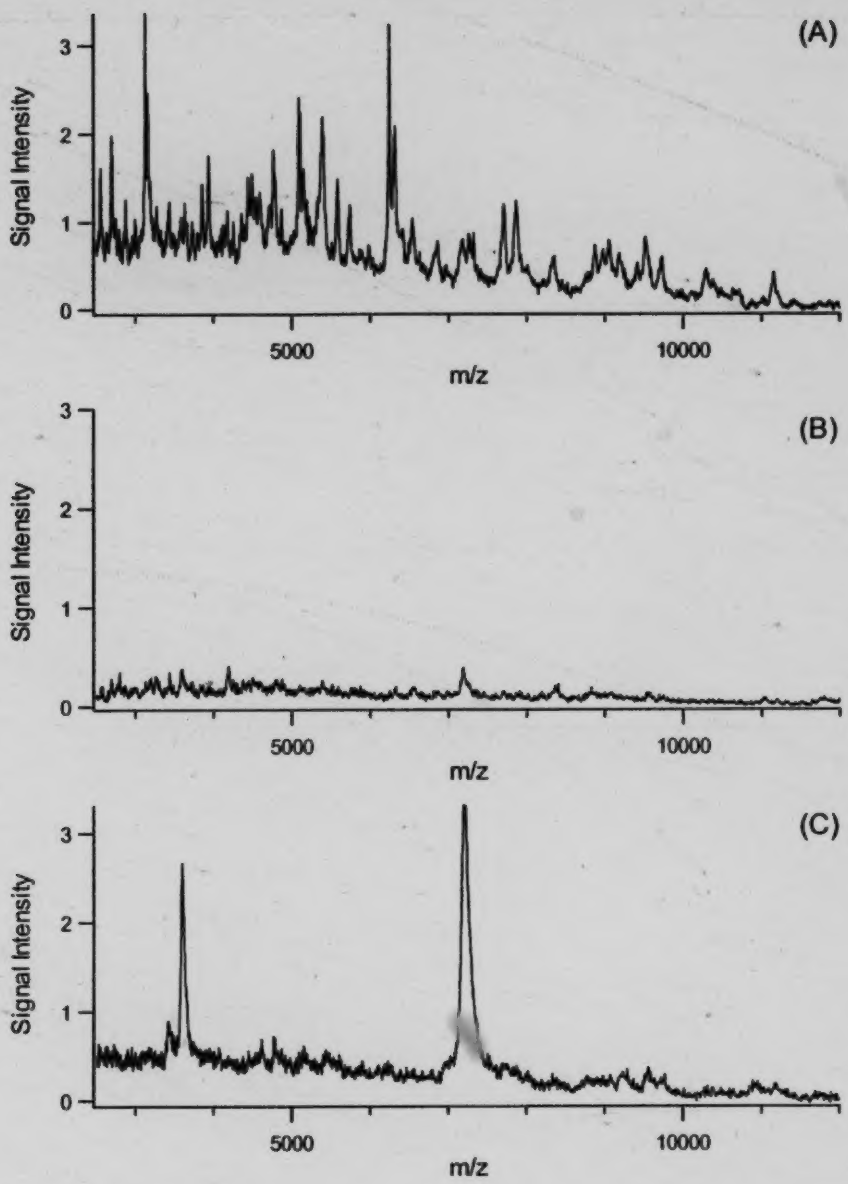

Figure 2.8 MALDI spectra obtained from sequential extractions of an *E. coli* sample: (A) first extract with 0.1% aqueous TFA, (B) second extract with 0.1% aqueous TFA, and (C) third extract with 0.2% SDS in 0.1% TFA.

This indicates that SDS can solubilize some proteins that were not solubilized by the 0.1% aqueous TFA solvent.

It is interesting to note that the peak shown in Figure 2.8C is unusually broad, suggesting that it is from a mixture of proteins such as those with different modifications. Tryptic digestion of this sample revealed several peptide ion peaks in the low mass region, but peptide mass fingerprinting alone was not sufficient to identify the protein(s). Unfortunately, ESI MS/MS could not be performed on this sample because of the presence of SDS. MALDI-MS/MS could potentially be used to identify the protein(s).

2.3.9 Protein Digestion. Limbach *et al.* has shown that the presence of SDS in a sample will change the relative intensity ratio of peptides having varied hydrophobicity and have used this finding to increase the number of detectable peaks following trypsin digestion [13,14]. In our work, we are interested in examining the effect of SDS on the digestion process. This may also hold some benefits, as it is expected that the added SDS aids in protein denaturation and therefore alters the digestion efficiency. This work is motivated by the fact that detection of a greater number of peptides and/or different types of peptides from a protein can enhance the confidence of protein identification by using either peptide mass fingerprinting or MS/MS in proteomics.

To study the effects of SDS on the digestion of proteins by trypsin, four different proteins (ubiquitin, cytochrome c, myoglobin and BSA) were prepared in either 0.1% or 1% SDS and digested with trypsin (see section 2.2.4 for details). As a control, the same protein samples were digested in the absence of SDS, and then SDS was added to the sample at concentration of 0.1% or 1%. The digests were then analyzed by MALDI-MS as described in section 2.2.4.

The MALDI spectra obtained from the digestion of 1 μg/μL ubiquitin are shown in Figures 2.9 and 2.10. Figure 2.9 corresponds to the control in which SDS was added to the sample after the digestion was complete. At all levels of SDS, a MALDI spectrum could be obtained with detectable ubiquitin peptide fragments. However, there are some notable differences between the spectra. The addition of SDS to the sample following digestion had an effect on the relative intensities of the peaks. This is particularly noticeable for the peaks corresponding to sequence coverage 34-42, 64-72, and 55-63. However, the addition of SDS resulted in few changes to the overall number of detected peaks in each spectrum. From Figure 2.9A, we see that 10 peaks were identified as peptide fragments of ubiquitin. When 0.1% SDS was added to the sample following digestion, the same 10 peaks were detected (Figure 2.9B) with an addition of a new peptide fragment corresponding to sequence 7-27. The addition of SDS to the digested sample also produced several peaks resulting from sodium attachment. The spectrum obtained from the sample containing 1% SDS is shown in Figure 2.9C. Eight of the 10 peaks originally detected in the sample without SDS were still visible, with no additional peptide fragments being observed. However, several low mass ions corresponding to SDS-matrix ion clusters were also observed in addition to sodium attachment peaks, thus deteriorating the quality of the MALDI spectrum.

Ubiquitin was then digested in the presence of SDS, and the resulting spectra are shown in Figure 2.10. With 0.1% SDS added to the sample prior to digestion (Figure 2.10B), seven additional peptide fragments were observed in the MALDI spectrum that were not detected in the sample digested without SDS or in the sample with 0.1% SDS added after digestion. This result indicates that the presence of SDS in the sample has an effect on the digestion process itself, rather than exclusively on MS detection.

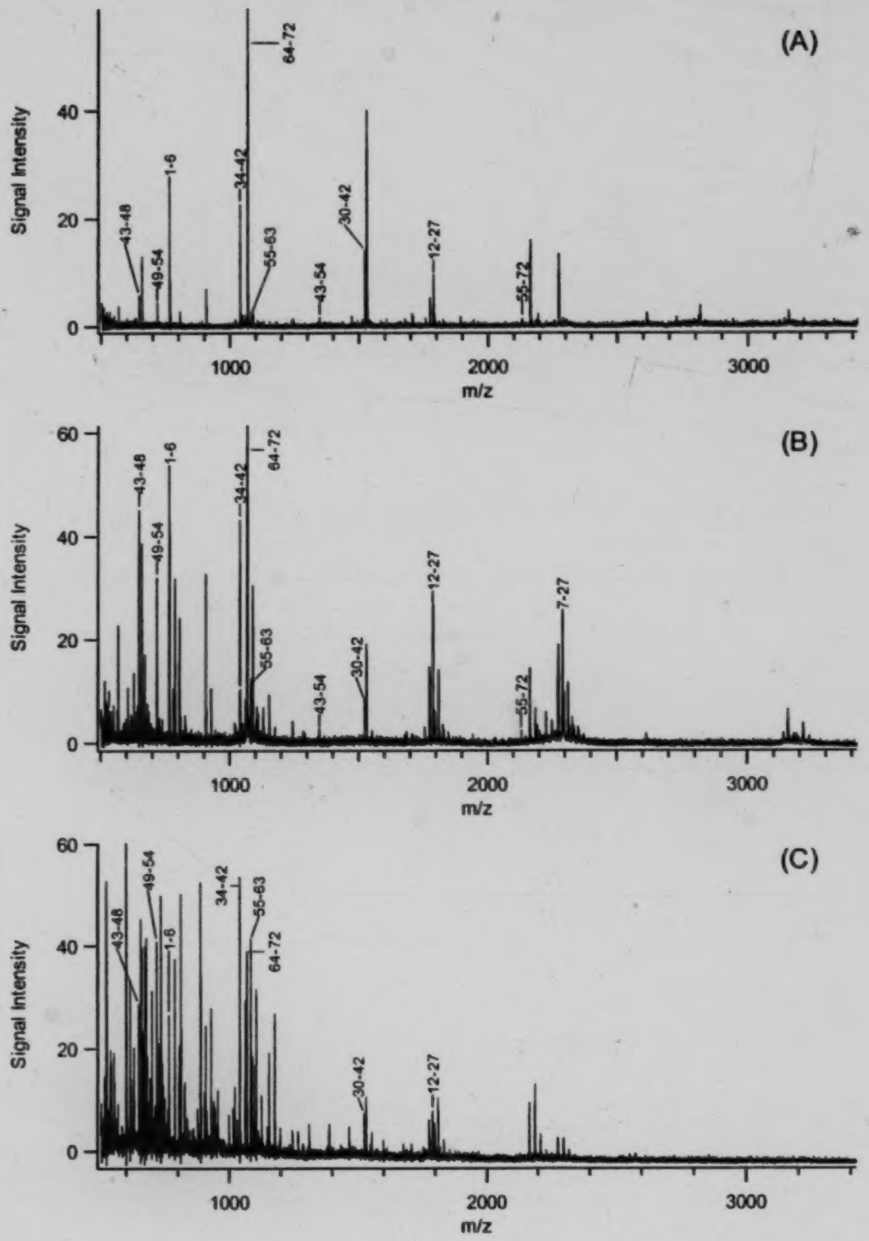

Figure 2.9 MALDI spectra obtained from digestion of 1 μg/μL ubiquitin (A) without adding SDS, (B) with 0.1% SDS added after digestion and (C) with 1% SDS added after digestion. The peptide fragments are labeled according to the amino acid sequence coverage of ubiquitin.

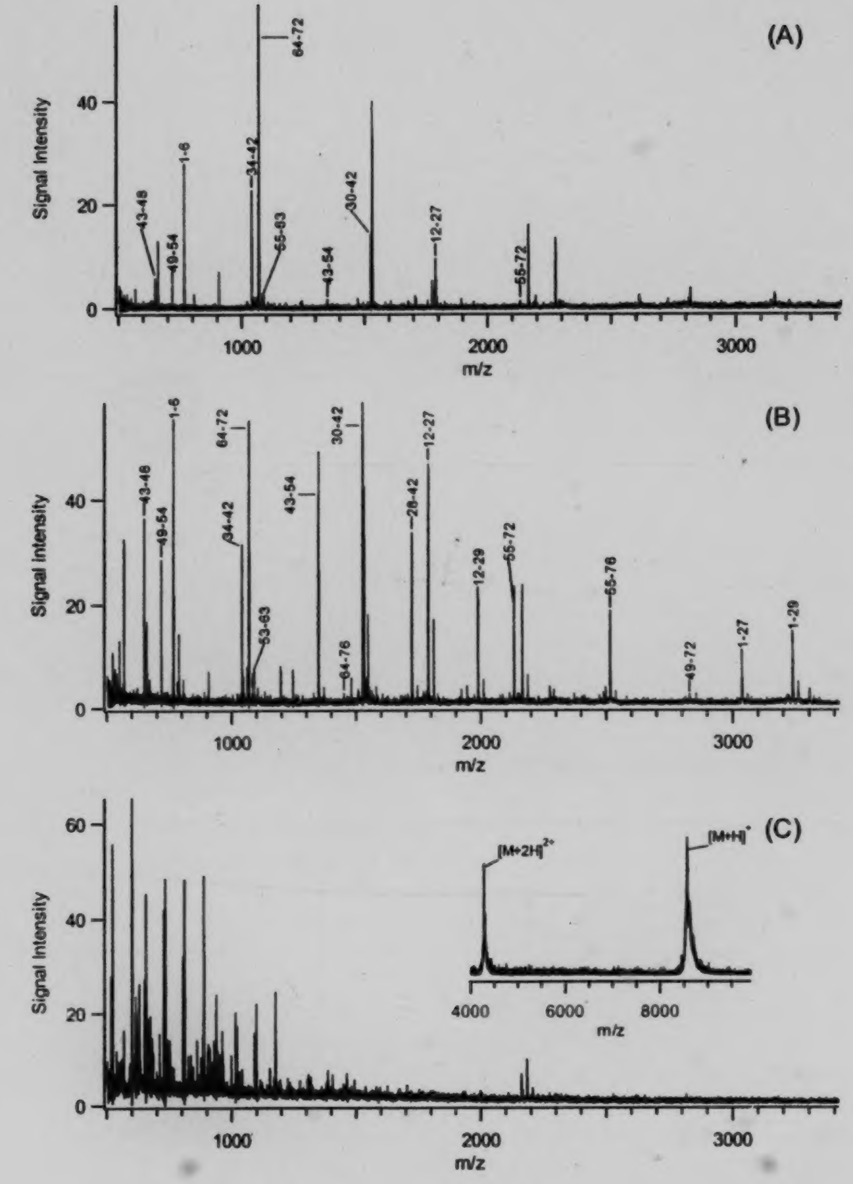

Figure 2.10 MALDI spectra obtained from digestion of 1 μg/μL ubiquitin (A) without adding SDS, (B) with 0.1% SDS added prior to digestion and (C) with 1% SDS added prior to digestion. The peptide fragments are labeled according to the amino acid sequence coverage of ubiquitin.

Although additional peaks were detected, no new cleavage sites of ubiquitin were observed. The additional peaks all contained missed cleavage sites, indicating that the addition of SDS perhaps somewhat altered the trypsin digestion efficiency. This is not surprising considering that SDS is a strong denaturing reagent and can cause the change of trypsin conformers. At the 1% level of SDS (Figure 2.10C), we see that no peptide fragments of ubiquitin were observed in the low mass region of the spectrum. However, the strong intact molecular ion was still observed, indicating that the sample remained undigested. Since peptides were observed in the control spectrum with 1% SDS added following the digestion, the results indicate that 1% SDS completely disrupts the activity of the trypsin, thereby preventing digestion. This is likely due to denaturation of the enzyme in the SDS containing solution.

Similar results were observed for the digestion of cytochrome c. However, when the digests of myoglobin and BSA were analyzed, only minor differences were observed between the control digests and the samples from digestion of proteins in the presence of SDS. Therefore it can be concluded that although the presence of SDS in the sample will change the digestion process, the extent of this change and the observed effect on the MALDI spectrum is dependent on the type of protein. Nevertheless, the results shown here demonstrate that proteins can be digested in the presence of low amounts of SDS, and that the resulting digests can be analyzed by MALDI-MS. The additional peptides produced are useful for protein identification by peptide mass mapping in proteome analysis. The additional peptides may also increase the chances of generating useful sequence information on a protein based on MS/MS.

2.4 Conclusions

We have demonstrated that a two-layer sample preparation method can be very useful in analyzing protein and peptide samples containing SDS by MALDI-MS. This method first involves the formation of a thin layer of matrix crystals, followed by depositing a mixture of sample and matrix dissolved in a proper solvent and with an adjustable sample to matrix ratio. Several variables related to the sample preparation are shown to affect the intensity of the MALDI signal. With this method, an on-probe washing step should be performed, if the deposited spot can tolerate this procedure. Commonly used solvents (*i.e.*, 50% acetonitrile/water or 40% methanol/water) for handling protein samples in MALDI can be utilized to prepare the second-layer matrix/analyte mixture. The matrix to SDS ratio is not critical, but the protein to SDS ratio can have a significant effect on the MALDI signal response. Thus dilution or concentration of the sample solution may be required when this method is applied to analyze samples with unknown protein concentration. The level of SDS should be kept at ≤ ~1%. With the two-layer method, there is no need of searching for hot spots in the process of acquiring MALDI spectra. If no analyte signals are observed in the first few shots, it usually indicates that the sample preparation conditions are not optimized and need to be adjusted.

The applications of the two-layer method for analyzing SDS-containing protein and peptide samples are illustrated. It was demonstrated that hydrophobic protein samples in the presence of SDS could be readily detected by MALDI. For direct analysis of cell extract by MALDI, it is now possible to incorporate solvent extraction involving SDS to examine more hydrophobic proteins. A more complete study on how SDS affects the digestion processes of various types of proteins, such as membrane proteins and glycoproteins, under various digestion conditions is currently underway [35]. Since the two-layer sample preparation method is tolerant of a certain percentage of SDS in peptide samples, we can readily analyze protein digests containing SDS. It is envisioned that sample preparation methods, such as the two-layer method shown here, that are capable of handling impure samples including those containing SDS, will greatly benefit the MALDI-MS and MS/MS approach in proteome identification.

2.5 Literature Cited

(1) Dalluge, J. J. *Fresenius J. Anal. Chem.* **2000**, *366*, 701-711.

(2) Patterson, S. D.; Aebersold, R. *Electrophoresis*, **1995**, *16*, 1791-1814.

(3) Rabilloud, T. *Electrophoresis*, **1996**, *17*, 813-829.

(4) Beavis, R. C.; Chait, B. T. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 6873-6877.

(5) Vorm, O.; Chait, B. T.; Roepstorff, P. *Proceedings 41st ASMS Conference on Mass Spectrom. Allied Topics*, San Francisco, CA, May 31-June 4, **1993**; P621.

(6) Rosinke, B.; Strupat, K.; Hillenkamp, F.; Rosenbusch, J.; Dencher, N.; Kruger, U.; Galla, H. *J. Mass Spectrom.* **1995**, *30*, 1462-1468.

(7) Schnaible, V.; Michels, J.; Zeth, K.; Freigang, J.; Welte, W.; Bühler, S.; Glocker, M. O.; Przybylski, M. *Int. J. Mass Spectrom. Ion Processes* **1997**, *169/170*, 165-177.

(8) Jeannot, M. A.; Jing, Z.; Li, L. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 512-520.

(9) Puchades, M.; Westman, A.; Blennow, K.; Davidsson, P. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 344-349.

(10) Galvani, M.; Hamdan, M. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 721-723.

(11) Amado, F. M.; Santana-Marques, M. G.; Ferrer-Correia, A. J.; Tomer, K. B. *Anal. Chem.* **1997**, *69*, 1102-1106.

(12) Breaux, G. A.; Green-Church, K. B.; France, A.; Limbach, P. A. *Anal. Chem.* **2000**, *72*, 1169-1174.

(13) Tummala, R.; Ballard, L. M.; Breaux, G. A.; Green-Church, K. B.; Limbach, P. A. *Proceedings 48th Conference of the American Society for Mass Spectrometry and Allied Topics*, Long Beach, CA, June 11-15, **2000**; P481-483.

(14) Saunders, C. J.; Frankel, L. K.; Green-Church, K. B.; Tummala, R.; Bricker, T.; Limbach, P. A. *Proceedings 48th Conference of the American Society for Mass Spectrometry and Allied Topics*, Long Beach, CA, June 11-15, **2000**; P489-P490.

(15) Jensen, C.; Haebel, S.; Andersen, S. O.; Roepstorff, P. *Int. J. Mass Spectrom. Ion Processes* **1997**, *160*, 339-356.

(16) Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299-2301.

(17) Weinberger, S. R.; Boernsen, K. O.; Finchy, J. W.; Robertson, V.; Musselman, B. D. in *Proceedings 41st ASMS Conference on Mass Spectrom. Allied Topics;* San Francisco, CA, May 31-June 4, **1993**, pp. 775a-b.

(18) Xiang, F.; Beavis, R. C. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 199-204.

(19) Xiang, F.; Beavis, R. C. *Org. Mass Spectrom.* **1993**, *28*, 1424-1429.

(20) Mock, K. K.; Sutton, C.W.; Cottrell, J. S. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 233-238.

(21) Bai, J.; Liu, Y. H.; Cain, T. C.; Lubman, D. M. *Anal. Chem.* **1994**, *66*, 3423-3430.

(22) Köchling, H. J.; Biemann, K. in *Proceedings 43rd ASMS Conference on Mass Spectrom. Allied Topics;* Atlanta, GA, May 21-26, **1995**, p. 1225.

(23) Hensel, R. R.; King, R.; Owens, K. G. in *Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics;* Atlanta, Georgia, May 21-26, **1995**, p. 947.

(24) Vorm, O.; Roepstorff, P.; Mann, M. *Anal. Chem.* **1994**, *66*, 3281-3287.

(25) Edmondson, R. D.; Campo, K. K.; Russell, D. H. in *Proceedings 43rd ASMS Conference on Mass Spectrom. Allied Topics;* Atlanta, Georgia, May 21- 26, **1995,** p. 1246.

(26) Li, L.; Golding, R. E.; Whittal, R. M. *J. Am. Chem. Soc.* **1996,** *118*, 11662-11663.

(27) Kussmann, M.; Nordhoff, E.; Rahbek-Nielsen, H.; Haebel, S.; Rossel-Larsen, M.; Jakobsen, L.; Gobom, J.; Mirgorodskays, E.; Kroll-Kristensen, A.; Palm, L.; Roepstorff, P. *J. Mass Spectrom.* **1997**, *32*, 593-601.

(28) Dai, Y. Q.; Whittal, R. M.; Li, L. *Anal. Chem.* **1996,** *68*, 2721-2725.

(29) Dai, Y. Q.; Whittal, R. M.; Li, L. *Anal. Chem.* **1999**, *71*, 1087-1091.

(30) Wang, Z.; Russon, L.; Li, L.; Roser, D. C.; Long, S. R. *Rapid Commun. Mass Spectrom.* **1998,** *12*, 456-464.

(31) Dai, Y. Q.; Li, L.; Roser, D. C.; Long, S. R. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 73-78.

(32) Juhasz, P.; Costello, C. E.; Biemann, K. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 399-409.

(33) Keller, B. O.; Li, L. manuscript in preparation.

(34) Barnidge, D. R.; Dratz, E. A.; Jesaitis, A. J.; Sunner, J. *Anal. Biochem.* **1999**, *269*, 1-9.

(35) Zhang, N.; Li, L. manuscript in preparation.

Chapter 3

Protein Concentration and Enzyme Digestion on Micro-Beads for MALDI-TOF Peptide Mass Mapping of Proteins from Dilute Solutions [3, 4]

3.1 Introduction

Since its introduction in the early 1990's, peptide mass mapping with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) has become a major tool for protein identification in proteomics. Detection by MALDI-TOF-MS has proven to be highly sensitive for the analysis of peptides resulting from protein digestion. In many cases, the detection limit of this technique can easily reach low femtomole quantities of analyzed sample. However, the successful generation of peptide fragments through protease cleavage from extremely low starting concentrations of protein remains problematic, and thus limits the ultimate sensitivity of the peptide mass mapping approach for protein identification. Specifically, digestion of a protein present at sub-micromolar concentrations is very difficult. As the protein concentration is reduced, the kinetics of the degradation reaction are reduced to the point where

---

[3] A form of this chapter has been published as: A. Doucette, D. Craft, and L. Li "Protein Concentration and Enzyme Digestion on Microbeads for MALDI-TOF peptide mass mapping of proteins from Dilute Solutions" *Anal. Chem.* **2000**, *72*, 3355-**3362**.

[4] D. Craft made significant contributions towards the development and optimization of the bead digestion procedure described in this chapter.

impractical digestion times are required. A simple approach to possibly compensate for the low reaction rate achieved using low substrate (protein) concentrations would be to increase the concentration of the proteolytic enzyme. However, the presence of a high enzyme-to-substrate ratio often results in a high degree of enzyme autolysis. These enzyme fragments can easily suppress the signals of the actual sample peptides in the MALDI spectrum, thus limiting the applicability of this approach for digestion of dilute samples.

Perhaps a better approach to the digestion of dilute proteins would involve the application of a method for concentrating the protein sample prior to enzyme digestion. However, many traditional concentration procedures, such as solvent evaporation, dialysis followed by solvent evaporation or the use of molecular weight cutoff filters, are generally time consuming. Of increased concern is that these preconcentration methods often result in a high degree of sample loss. In addition, with these methods, contaminants such as non-volatile salts, detergents and other agents present in the sample may also be concentrated along with the protein. In some cases, the loss in signal intensity from increased contaminant concentration may outweigh the benefits of enhanced protein concentration. Contaminants become a larger concern as we approach the detection limits of the instrument, since the MALDI technique is less tolerant of contaminants at lower sample concentrations [1].

An alternative approach is to *selectively* concentrate the proteins from a dilute, contaminated solution. This technique typically involves capturing the protein through interactions with a solid support having an affinity for the protein. To this end, commercially available cartridges and membranes have been developed that rely on hydrophobic interaction, antibody affinity-capture, or ion exchange, to selectively capture proteins (and not the contaminants) from dilute solutions. Several groups have reported the use of immobilized beads or surfaces with various affinity properties to selectively concentrate one type or a group of proteins [2-13]. However, the focus of these studies was not to concentrate proteins for subsequent elution and digestion. Protein elution from these capture devices typically involves the use of high percentages of organic solvents, which must again be reduced prior to enzymatic digestion.

In a related work, Vandekerckhove *et al.* demonstrated that peptides could be concentrated from solution using reversed-phase chromatography beads [14,15]. The beads were placed directly into the solution and the peptides were allowed to adsorb to the bead surface. The peptides were then directly analyzed by placing the beads on a MALDI target where they were mixed with a matrix solution. This method proved to be very effective at concentrating peptides. However, the success of this technique is ultimately limited by the sample. The increased affinity of hydrophobic peptides, or even larger peptides for the hydrophobic surface may prevent these samples from being detected. In this case, these peptides will not effectively mix with the matrix as they will remain tightly bound to the support.

An interesting strategy for protein digestion following capture on a solid support has been reported in the literature. Aguilar *et al.* reported the use of micro-beads for protein capture, followed directly by enzyme digestion of the adsorbed proteins [16]. They demonstrated that proteins could be digested while bound to a C_4 or C_{18} sorbent. The authors reported on the digestion of cytochrome c with trypsin and elution of the resulting peptide fragments from the sorbent for separation and detection by HPLC with UV absorbance detection. They reported that the digestion of cytochrome c was partially hindered due to protein adsorption to the sorbent, and also that following digestion, some of the generated peptides eluted from the sorbent into the solution phase. We note that strong affinity interactions can prevent trypsin digestion of bound proteins, as reported on the study of combining avidin-biotin chemistry with MALDI-TOF [7]. In their method, the authors used a high concentration of cytochrome c in order to fully saturate the solid support, and reported that digestion was completely hindered if the support was not fully saturated with cytochrome c. Thus, based on these results, this method would not seem applicable to dilute protein samples.

An alternative approach is to use immobilized trypsin to digest dilute protein samples [17-24]. Using this technique, impressive results were reported with protein concentrations of ≤100 nM [22-24]. However, in our lab it was found that at these concentrations, trypsin autolysis is still a problem for highly salt-contaminated samples [25]. The autolysis products are likely from unbound trypsin that is either present in the original immobilized trypsin sample or produced during storage or digestion.

In this chapter, a method is presented for the concentration and digestion of proteins from dilute, contaminated samples by adsorption of the protein on polymeric micro-beads. A washing step was incorporated in the procedure in order to remove contaminants from the sample prior to trypsin digestion. The resulting digests were directly analyzed by MALDI-TOF-MS through deposition of the polymeric beads on the MALDI target. The results of these experiments reveal that, with the micro-bead preconcentration procedure, very complete peptide mass maps can be routinely generated from highly contaminated protein samples present at concentrations of 100 nM or below.

3.2 Experimental

3.2.1 Chemicals and Reagents. Bovine serum albumin (BSA), equine cytochrome c, chicken lysozyme, and trypsin (from bovine pancreas, TPCK treated to reduce chymotrypsin activity; dialyzed, lyophilized) were from Sigma Aldrich Canada (Oakville, ON). 20 μm POROS R2 beads were from PerSeptive Biosystems (a gift of Professor F. Cantwell, University of Alberta). Prior to use, the beads were twice washed with water and suspended in methanol at a concentration of 10 mg/mL. Analytical grade acetone, methanol, acetonitrile, acetic acid, and trifluoroacetic acid (TFA) were purchased from

Caledon Laboratories (Edmonton, AB). Water used in the experiments was from a NANOpure water system (Barnstead/Thermolyne). All other chemicals were purchased from Sigma Aldrich Canada.

3.2.2 Protein Adsorption onto Beads. 100 nM protein samples of BSA, lysozyme, and cytochrome c were prepared in water and 100 μL aliquots were placed into separate siliconized vials. Before the addition of beads, the lysozyme sample was reduced to break disulfide bonds using 1 μL of 90 mM dithiothreitol (DTT), with incubation at 56°C for 30 minutes. The cysteines were then blocked to prevent reformation of disulfide bonds using 1 μL of 200 mM iodoacetamide, with incubation at room temperature in the dark for 30 minutes. To the protein solutions, 4 μL of the bead suspension (10 mg/mL in methanol) were added. The vials were then agitated with a vortex mixer for 10 minutes to allow for protein adsorption onto the beads. The beads were then pelleted in a centrifuge, and the supernatant was removed with a fine pipette tip. The beads were washed with three changes of water (100 μL per wash), agitating and pelleting the beads as described previously, before removing the supernatant. The protein samples, now bound to the R2 beads, were then digested with trypsin.

3.2.3 Enzyme Digests of Surface-Bound Protein. To the beads containing the adsorbed protein, 5 μL of 100 mM NH_4HCO_3 and 5 μL of 62 ng/μL trypsin solution (pH ~8) were added. The vials were vortexed briefly to suspend the beads and the resulting suspensions were incubated for 30 minutes at 37°C. Following the digestion, a final wash was performed by adding 100 μL of water to each bead sample, followed by vortexing and pelleting of the beads. The supernatant was removed from each vial, and 6 μL of water was added to the beads. The resulting samples were stored on ice prior to MALDI analysis.

3.2.4 Enzyme Digests of Protein in Solution. The three protein standards were also digested in free solution, without any preconcentration or cleanup step. 20 μL aliquots of the protein solutions, at 500 nM, 200 nM, or 100 nM were mixed in siliconized vials with 2 μL of 1 M NH_4HCO_3 and 1 μL of 0.1 μg/μL trypsin. In the case of lysozyme, the sample was again treated with DTT and iodoacetamide prior to tryptic digestion. The samples were incubated at 37°C for times ranging from 30 minutes to 2 hours, and then a portion of the digested sample was mixed with a saturated matrix solution. These samples were also stored on ice prior to MALDI analysis.

3.2.5 MALDI Sample Preparation. The matrix used in these experiments was α-cyano-4-hydroxycinnamic acid (HCCA). The HCCA, purchased from Aldrich, was first purified by recrystallization from ethanol. For the samples that were digested in free solution (no beads added), a two-layer MALDI deposition method was used [26]. This involves the deposition on the MALDI target of a microcrystalline matrix layer via fast evaporation from a 1 μL solution of HCCA dissolved in 80% acetone/methanol at a concentration of 12 mg/ml. 1 μL of the digested protein sample was mixed with either 2 or 4 μL of a saturated HCCA solution in 40% methanol in 0.2% aqueous TFA. The protein-matrix solution was briefly vortexed and a 0.4 μL portion was deposited on top of the first matrix layer. Once dried, the spot was washed three times with 0.75 μL of room temperature water.

For the samples bound to the R2 beads, a 3 μL portion of the bead slurry, corresponding to approximately half the total available sample, was deposited on the probe tip. This was allowed to evaporate to complete dryness before a 0.75 µL portion of saturated HCCA in 50% acetonitrile/0.2% TFA in water was deposited on top of the dried beads. This was allowed to dry in a slow stream of air, which aided in the distribution of the beads over the surface of the probe tip (see section 3.3.4). No on-probe washing step was performed for the bead samples.

3.2.6 Instrumentation. MALDI spectra were collected on a linear time-lag focusing MALDI time-of-flight mass spectrometer, which was constructed at the University of Alberta and has been described in detail elsewhere [27]. A pulsed nitrogen laser operating at 337 nm was used to generate the MALDI ions. The laser was directed at a portion of the target containing the beads. The first 5 to 10 shots from a new spot were discarded and the next 10 to 200 shots were recorded. For the purpose of comparing resolution and mass accuracy, spectra were also obtained on a PerSeptive Biosystems Voyager DE Elite reflectron MALDI time-of-flight mass spectrometer. All spectra were the results of signal averaging of between 100 and 200 shots. Data processing was done with the IGOR Pro software package (Wavemetrics Inc., Lake Oswego, OR).

3.3 Results and Discussion

The procedure reported herein involves three distinct steps. In step 1, the protein sample is concentrated and purified by adsorption from solution onto a solid support, which is followed by a washing step to reduce contaminants. In step 2, the protein sample is digested while bound to the support. The peptides bound to the beads are then washed. In step 3, the beads are loaded to the MALDI-TOF mass spectrometer for detection. Using this procedure, a dilute, contaminated protein sample can be effectively concentrated prior to digestion. Therefore, the difficulties that are normally associated with digestion of very dilute samples are greatly reduced. While the concept of this overall experimental scheme is very simple, each step had to be optimized to achieve the desired performance. The method development and performance is described below.

3.3.1 POROS R2 Beads. The POROS R2 beads used in these experiments were poly(styrene divinyl benzene) reversed-phase chromatography beads having a retention property similar to conventional C_8 or C_{18} supports. It has already been shown by Vandekerckhove *et al.* that small peptides (below approximately 2000 Da) will adsorb to these beads and that these peptides can be directly detected by placing the beads on a MALDI target [14]. This type of bead will also bind larger peptides and proteins, such as insulin chain B and BSA. However, it is suspected that these larger molecules may not be detected by MALDI-TOF MS with the same degree of sensitivity as the smaller peptides. This is due to the fact that the binding strengths of the molecules are higher than that of the smaller peptides, and thus they cannot be efficiently eluted from the bead surface in the MALDI process. Adsorption of proteins to these beads, with no selectivity for a particular class of protein, is a desirable feature for concentrating a wide range of samples. However, the interaction between protein and surface cannot be too strong, or it will interfere with the subsequent digestion (and detection) [7].

The amount of beads used in these experiments was optimized to produce the best MALDI spectra following digestion of the samples, and was also easy to handle. For most experiments reported in this work, 40 μg of R2 beads, suspended in 4 μL of methanol, were added to 100 μL of 100 nM protein solutions. At lower protein concentrations, the amount of beads was adjusted to lower amounts.

3.3.2 Optimization of Trypsin Digestion. The conditions for the trypsin digestion of proteins bound to beads were optimized using cytochrome c at a concentration of 100 nM. 40 μg of POROS R2 beads were added to 100 μL aliquots of the cytochrome c and the mixtures were shaken for 10 minutes on a vortex mixer. Once the beads were pelleted, the aqueous layer could be removed with a fine pipette tip with minimal loss of the beads. If we assume that essentially all of the cytochrome c is present on the beads, this would represent a total of 10 pmol, or 124 ng of cytochrome c. Next, 10 μl solutions containing various concentrations of trypsin in 50 mM NH_4HCO_3 were added to the bead samples and the suspensions were incubated at 37°C for various times before analyzing the digested peptides adsorbed to the beads by MALDI-TOF mass spectrometry. It was found that digestions using 310 ng of trypsin produced the best spectra, in the shortest amount of time. When a 1:1 mass ratio of trypsin to cytochrome c was used, very few digestion products were detected after a 1-hour digestion period. Adding more trypsin (10:1 by mass) revealed too many trypsin autolysis peaks in the MALDI spectra, thus reducing the spectral quality. Also, it was found that for 310 ng of trypsin, a 30-minute digestion period was sufficient to produce several intense peptide fragments of cytochrome c.

3.3.3 Sample Deposition and Detection by MALDI-TOF-MS. In our experiments, following trypsin digestion, the beads were directly placed on the sample target for MALDI analysis. It was found that after the protein samples were digested, the resulting peptide fragments would remain bound to the beads rather than elute into aqueous solution. These results were as expected as it has already been shown that small peptides will adsorb on the surface of this type of bead [14]. Evidence for peptide adsorption is also given by the fact that the peptide fragments were still detected on the beads by

MALDI-TOF-MS with no loss in signal intensity, even after the beads were extensively washed following trypsin digestion.

Several attempts were made to elute the peptides from the beads in a vial by adding a saturated HCCA solution containing various percentages of acetonitrile in 0.1% TFA, but the spectral quality obtained from these samples was lower than that obtained by placing the beads directly on the target. In addition, this approach proved to be extremely difficult to maneuver, since it was desirable to use only a small volume of matrix solution for elution of the peptides from the beads in order to prevent dilution. With such a small volume, it is difficult to transfer matrix solution to the MALDI target without also transferring several beads. It was therefore concluded that placing the beads directly on the target, followed by elution of the peptides with matrix solution provided the simplest approach for MALDI-TOF-MS analysis.

When we initially began these experiments, an aliquot containing the equivalent of approximately 4 μg of R2 beads was placed on the MALDI target. After allowing the beads to dry, 0.75 μL of a saturated HCCA solution in 50% acetonitrile/0.2% aqueous TFA was placed on the beads and allowed to evaporate to dryness. It was observed that as the acetonitrile evaporated, the beads drew together on the target, and a uniform layer of HCCA crystals formed over the spot. It was found that the strongest signals were obtained when the laser was aimed at a region of the target containing the beads, rather than at a spot that only contained matrix crystals. If the beads were blown off the target with a stream of air, leaving only the matrix, only weak signals could be observed from any spot on the target. We therefore added more beads to a single target, placing an aliquot containing ~20 μg of beads on the MALDI probe tip. Also, after placing the matrix solution on top of the dried beads, we allowed the bead-matrix suspension to dry under a stream of air, which created a more uniform layer of beads over the probe tip (see Figure 3.1). It was found that directing the laser at a region of the target containing multiple layers of the beads resulted in either very weak signals, or no detectable signals in the MALDI spectrum. Drying the beads under a stream of air reduced the multiple bead layer formation, creating a more uniform sample deposition, and therefore resulted in increased signal intensity as well as reproducibility over the entire sample spot.

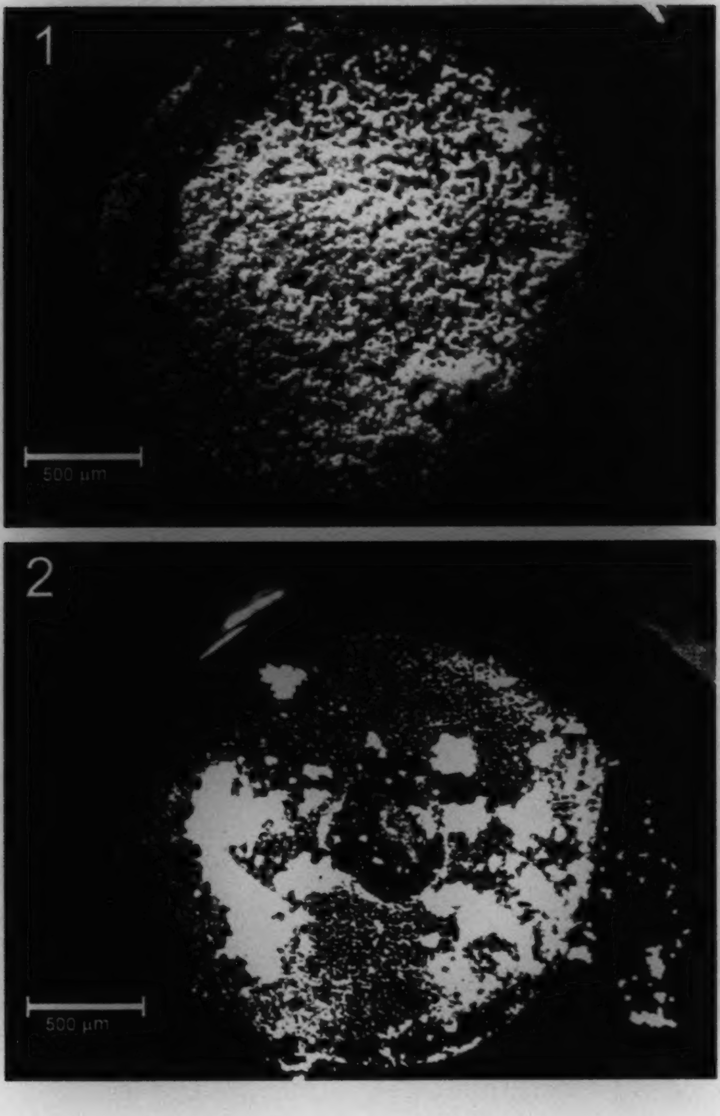

Figure 3.1 Distributions of micro-beads on a MALDI probe tip. About 20 µg of POROS R2 beads (20 µm diameter) were spotted on the probe tip following adsorption and digestion of a protein on the beads. Visible in the pictures are the R2 beads, shown as the large spheres, as well as the matrix crystals, shown as smaller dots. (A) The matrix solution was allowed to dry under a stream of air, which distributes the beads more evenly over the target. (B) The sample was undisturbed as the matrix solution evaporated. This caused the beads to draw together, forming non-uniform multiple layers.

3.3.4 Mechanism of Peptide Desorption From MALDI Target. The experimental results discussed above suggest that, even after the matrix solution is added to the beads, most of the peptide fragments would ultimately remain adsorbed to the polymeric bead surface. It is theorized that when the matrix solution is added to the beads on the target some peptides will elute, while for other peptides, only a portion of the molecule (*i.e.* the soluble domain) makes its way into the solution while the insoluble domain is still attached to the bead surface. The degree of solubilization is dependent on the solvent composition and peptide properties. As the acetonitrile in the matrix solution evaporates and the aqueous content of the matrix solution increases, the HCCA will begin to precipitate, and the matrix will co-crystallize with the fully eluted or partially bound peptides. When the laser is directed at the bead surface, the energy transferred will dissociate the binding interaction between the peptide and the bead surface and lift the peptide to the gas phase. This theory of peptide release from the bead surface by laser desorption was also presented by Papac *et al.* in a previous work [2]. As additional supporting evidence for this idea we observed that the laser threshold energy required to ionize the sample is slightly higher (~5%) than for a sample at the same concentration on the target without beads, indicating partial binding to the bead surface [28].

3.3.5 Comparative Analysis of Uncontaminated Samples. Solutions of BSA, cytochrome c, and lysozyme at concentrations ranging from 500 nM to 100 nM were prepared. The conditions for the digestion of these proteins in free solution were optimized and the MALDI spectra were recorded. It was found that while the 500 nM protein samples produced reliable peptide mass maps for all samples, digestion products could not consistently be detected for the 100 nM samples. However, when the 100 nM protein samples were digested following preconcentration on R2 beads, the resulting peptide mass maps were of higher intensity, produced a greater number of peaks, and gave higher sequence coverage than the corresponding 500 nM samples digested without preconcentration. These results are compared more clearly in Figure 3.2.

Figure 3.2 displays MALDI spectra recorded from the digestion of lysozyme under different conditions. Figure 3.2A is the MALDI spectrum obtained from the analysis of a 500 nM solution of lysozyme that was digested with trypsin for 30 minutes. One can see from this spectrum that several peaks are visible which correspond to peptide fragments of lysozyme. In all, 12 peaks were detected, representing sequence coverage of 83%. Figure 3.2B shows the result of a digestion of 100-nM lysozyme in solution. As seen in this figure, no peptide fragments of lysozyme are detected. Only peaks corresponding to trypsin autolysis are seen in the spectrum. Figure 3.2C shows the MALDI spectrum of a 100 nM lysozyme solution that was digested on beads following preconcentration of the protein. A total of 26 peaks were assigned to lysozyme fragments, representing 100% sequence coverage. This compares favorably to the 12 peaks that were detected from the 500 nM solution without preconcentration. Also, the signal intensities are much higher, as indicated by the absolute intensities of the y-axis shown in Figures 3.2A-C. The signals are approximately 5 times more intense for the 100 nM lysozyme solution that was digested on beads than for the 500 nM sample without preconcentration.

The signal enhancement seen in Figure 3.2C is mainly due to the increased sample loading on the MALDI target that is achieved using the bead concentration procedure. In addition, digestion of 500-nM protein in a free solution might not be complete under the optimized conditions used for a 30-minute digestion. To digest a

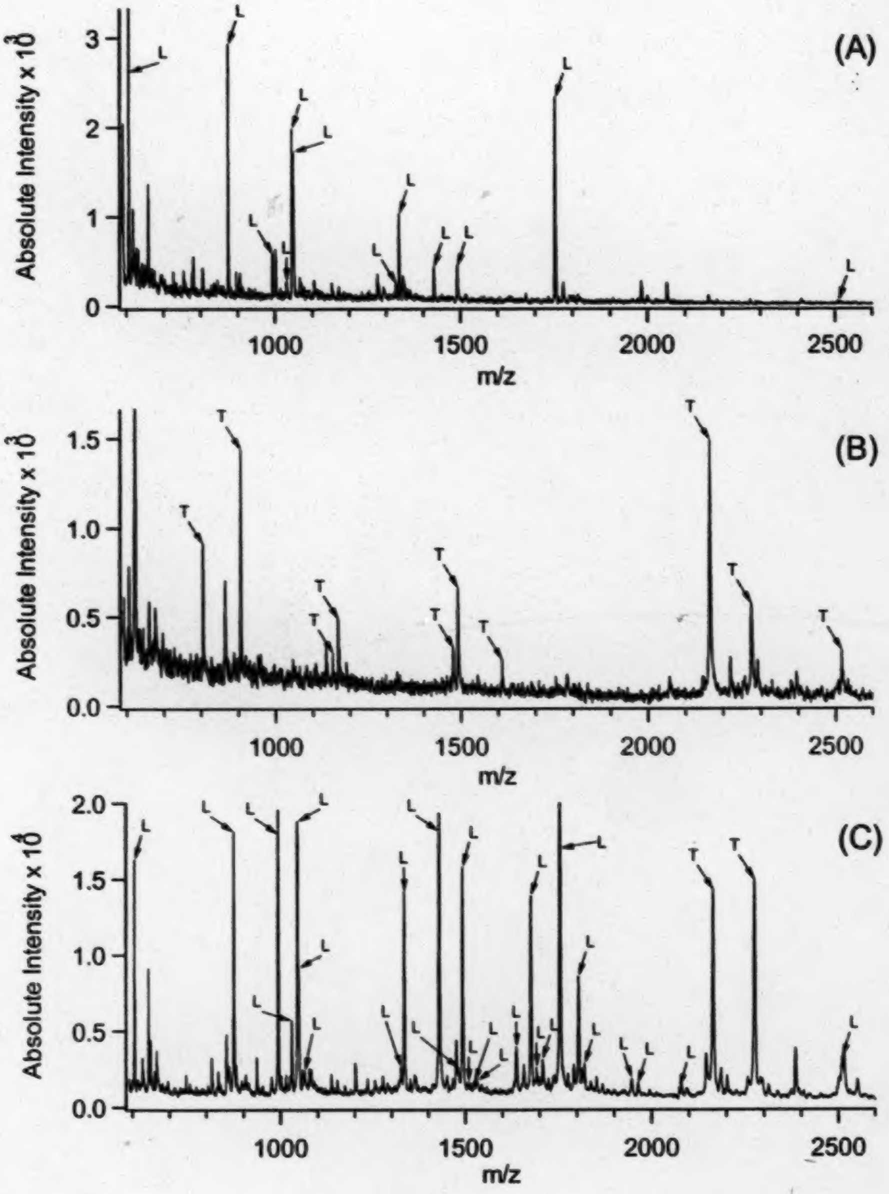

Figure 3.2 MALDI spectra obtained from the digestion of (A) 500 nM lysozyme in solution, (B) 100 nM lysozyme in solution and (C) 100 nM lysozyme that was concentrated and digested on R2 beads.

dilute protein in a free solution with minimum enzyme autolysis, the enzyme to protein ratio must be kept low. After digestion, peptides could be concentrated with devices such as the ZipTip (Millipore). It was found that, using a C_{18} Ziptip to concentrate the peptides following trypsin digestion of 100 nM protein solutions, strong signals comparable to those obtained with the bead concentration procedure were observed. The protein to enzyme ratio was optimized to be 0.8:1 for this Ziptip experiment. Under these conditions, the kinetics of the reaction remained low and the samples required incubation times on the order of 24 hours to achieve complete digestion. As protein concentration decreases further (e.g., 10 nM protein), the incubation time becomes impractical. Thus the use of micro-beads to pre-concentrate the protein has the potential advantage of shortening the digestion time as well as the possibility of handling lower concentrations of protein solutions, compared to the ZipTip method.

One of the major concerns we had prior to performing these micro-bead experiments was that the digestion of a protein while bound to a polymeric surface would be hindered or blocked. It was suspected that incomplete digestion, if any, would result since several cleavage sites would not be accessible to the trypsin while the protein was adsorbed to a solid support. This was observed by Aguilar *et al.* who demonstrated that for proteins adsorbed to C_4 or C_{18} sorbents, specific sites of the proteins were inaccessible to tryptic cleavage [16]. This concern also arose from our previous experience with biotinylated proteins that were complexed to avidin-coated beads. In this case, the complexed proteins do not undergo efficient digestion [7]. However, the chemistry of the polymeric beads used in this present work is quite different from the avidin-biotin chemistry. In addition, steric interference on the R2 beads should be much less than on avidin-bound beads. The protein binding to the R2 beads, which occurs through a non-specific binding process, should be much weaker than the avidin-biotin interaction. Also the differences in surface chemistry for polymeric beads versus C_4 or C_{18} silica may also lead to differences in enzyme cleavage efficiencies. The results shown here clearly indicate that the digestion of a protein is not hindered while adsorbed to the R2 bead surface. Since the type of beads can be different in size and surface chemistry, a detailed investigation of the effect of bead type on enzyme digestion efficiency is warranted. This is particular true in light of the different results on enzyme digestion efficiency found in this work (using polymeric beads) and that of reference 16 (using C_4 or C_{18} silica beads).

For the digestion of proteins on the R2 beads, the conformation of a protein adsorbed on this hydrophobic surface is expected to be quite different than in solution. It has been reported that protein molecules will achieve maximal binding by orienting on a hydrophobic surface such that the hydrophobic domains of the protein bind to the solid support, leaving the hydrophilic domains facing the solution phase [29-34]. This necessitates a change in protein conformation in order to expose hydrophobic domains, normally buried in the interior of the protein in solution, to the exterior of the protein [29-34]. We propose that the trypsin begins by digesting the hydrophilic domain of the protein, which is exposed to the solution phase. Once these segments are released by the cleavage process, other cleavage sites that were previously inaccessible to the enzyme will become exposed. The enzyme will then digest these sites. The final result would be a complete, or near complete, digestion of the protein.

3.3.6 Concentration Limit. The bead digestion procedure was applied to samples at lower concentrations. Figure 3.3 displays the MALDI spectrum obtained from the digestion of a 10 nM cytochrome c solution (500 μL total volume). A total of 40 μg of R2 beads was added to the sample vial for protein adsorption. Owing to the large sample volume of this dilute solution, a total of 5 pmol of protein is available for digestion. Thus, by capturing the cytochrome c into a small volume on the surface of the R2 beads, a sufficient concentration of protein is available for improved reaction kinetics, and high signal intensities in the MALDI spectrum following digestion.

From Figure 3.3, we can see that a 10 nM protein concentration was sufficient to produce a very complete peptide map for cytochrome c. A total of 29 peaks were detected, representing 97% sequence coverage. Although it was possible to digest samples at even lower concentrations, it was found that as the concentration was reduced even further, the quality of the MALDI spectra deteriorated significantly. Several peaks became visible in these spectra whose origin may be attributed to contaminating proteins such as keratins. These proteins, if present in the solution, will also concentrate by adsorption to the bead surface, and will therefore be digested, and detected by MALDI-TOF-MS.

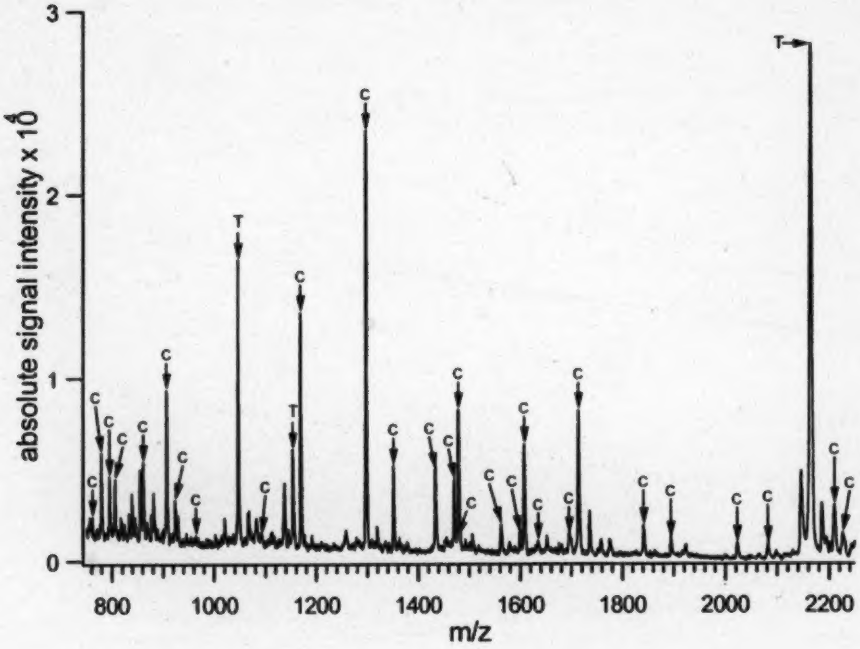

Figure 3.3 MALDI spectrum showing the trypsin digestion of 10 nM cytochrome c that was concentrated on R2 beads. 500 ml of sample were used and the digestion time was 30 minutes. A total of 29 peaks were assigned to peptide fragments of cytochrome c, resulting in 97% sequence coverage.

3.3.7 Analysis of Samples Contaminated with NaCl and Urea. Non-volatile salts such as NaCl and chaotropic agents such as urea are commonly used in protein sample work-up. To illustrate the performance of the micro-bead concentration procedure for analyzing protein samples containing these contaminants, solutions of BSA, cytochrome c, and lysozyme were prepared in 2 M NaCl or 2 M urea. 500 nM solutions were first digested with trypsin in free solution using the previously optimized conditions for the corresponding uncontaminated samples. Using the two-layer method to deposit the protein samples prepared in 2 M NaCl, it was found that after the sample had dried on the probe tip salt crystals were visible on the surface of the sample spot. The protein samples prepared in 2 M urea also produced a noticeable precipitate on top of the sample spot. An on-probe washing step was performed by adding water to the sample spot, waiting a few seconds, and removing the water with a Kimwipe. This procedure was repeated a total of three times per sample spot. After the washing step, it was observed that the remaining sample spot had "holes" in the matrix crystalline layer, resulting from regions where the crystallized salt or urea had dissolved during the washing step. However, by aiming the laser at a cleaner, more uniform area of the sample layer signals could be obtained.

Figure 3.4 displays the remarkable ability of the MALDI technique to ionize samples prepared in the presence of contaminants such as NaCl and urea. The spectra of Figure 3.4 are for digestion of the salt and urea contaminated samples of BSA, as well as the digestion of a clean BSA sample. Figure 3.4A shows that when the 500 nM BSA sample prepared in pure water was digested with trypsin in free solution several peptide fragments are visible in the MALDI spectrum with good signal intensities. However, when 2 M NaCl was present in the sample the quality of the resulting spectrum was greatly reduced, as seen in Figure 3.4B. However, 6 of the 21 peaks detected in the clean sample were visible, although the intensity of these signals dropped considerably. Figure 3.4C displays the spectrum obtained for the sample contaminated with 2 M urea, which had a less pronounced effect on the spectrum than 2 M NaCl. However, the signal intensities dropped by over half as compared to the uncontaminated BSA sample, and only 11 of the 21 peaks originally detected were observed. It should be noted that by

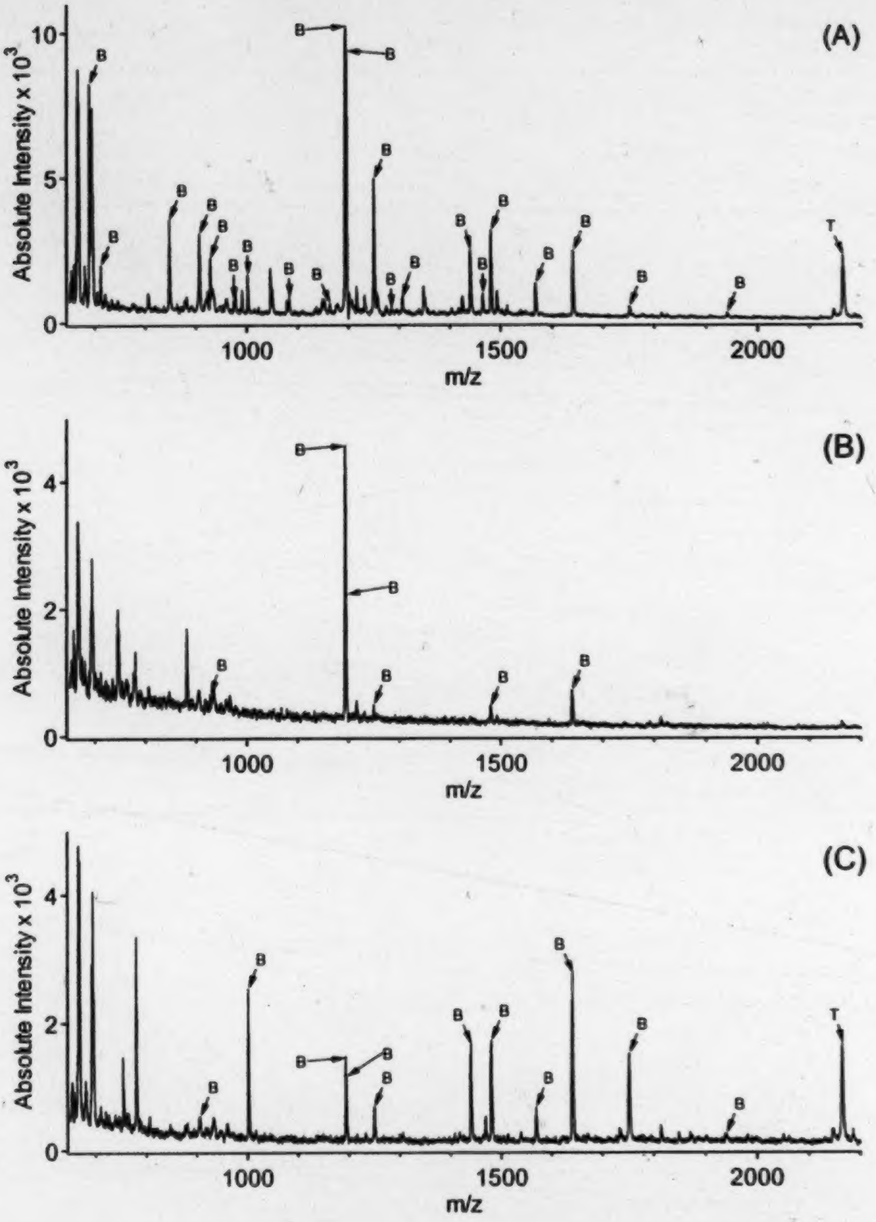

Figure 3.4 MALDI spectra obtained from the free solution digestion of 500 nM BSA in (A) pure water, (B) 2 M NaCl and (C) 2 M urea. An on-probe washing procedure was incorporated for the analysis.

omitting the on-probe washing step, no peptide signals could be detected from MALDI-TOF-MS analysis. Figure 3.4 clearly illustrates that, even with the on-probe washing step, a high concentration of NaCl and urea can affect the overall performance of the MALDI analysis.

Next, 100 nM solutions of the proteins in either 2 M NaCl or 2 M urea were digested using the bead concentration procedure. The beads were washed three times prior to digestion and once following the 30-minute incubation period with trypsin. The peptide mass maps resulting from the digestion of the protein samples were obtained.

Figure 3.5 shows the MALDI spectra obtained from the 100 nM BSA samples digested on beads. When 100 nM BSA in 2 M NaCl is digested on beads, following the washing steps, we see that the salt has very little effect on the quality of the MALDI spectrum (Figure 3.5B). The intensity of the signals is essentially constant and only 2 of the 28 peaks are missing when compared to the clean BSA digest (Figure 3.5A). The presence of 2 M urea only had a small effect on the quality of the spectrum, as shown by comparison of Figure 3.5C and 3.5A. The signal intensity in Figure 3.5C has dropped somewhat, however 25 of the 28 peaks are still detectable.

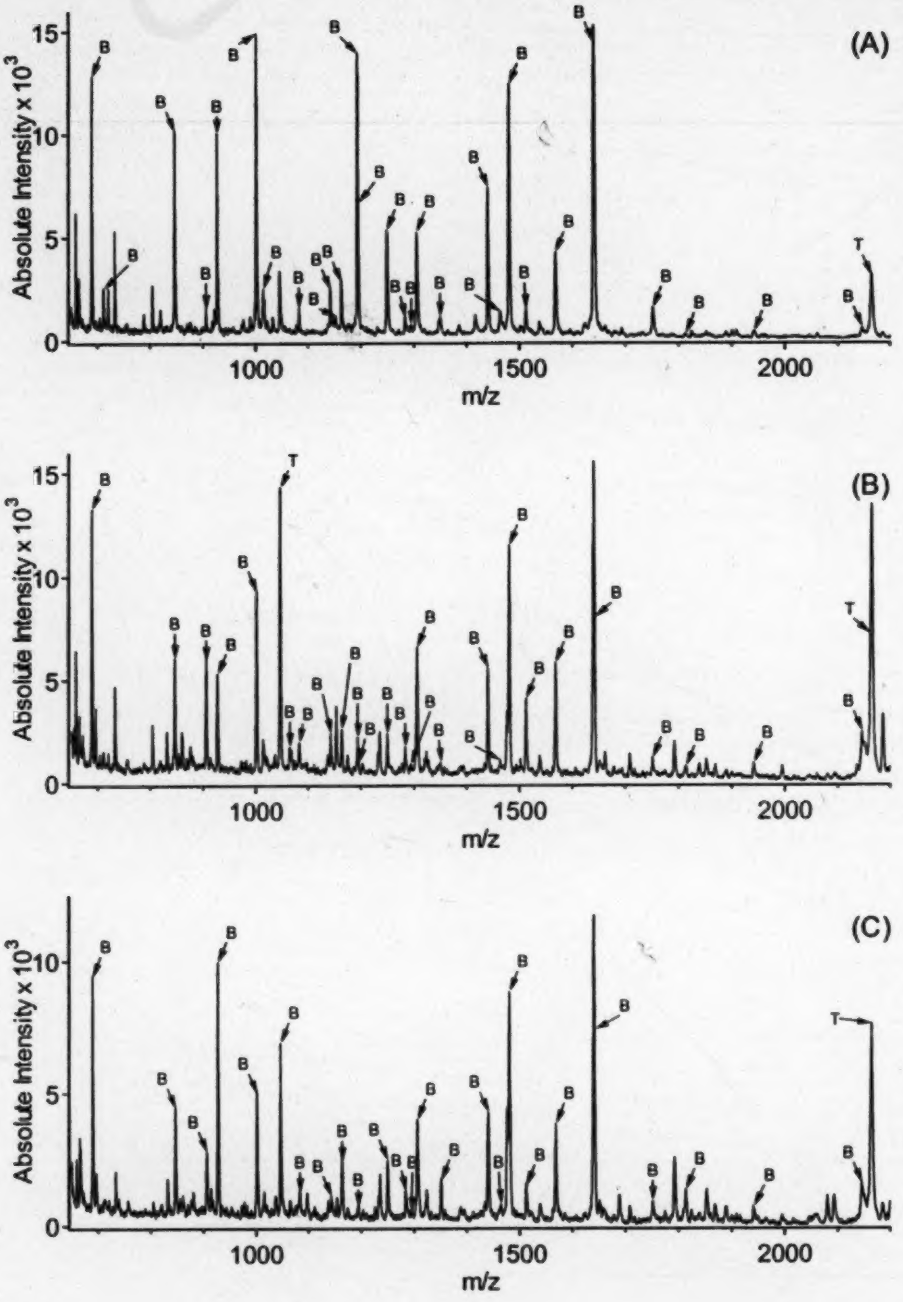

Figure 3.5 MALDI spectra obtained from micro-bead preconcentration and digestion of 100-nM BSA in (A) pure water, (B) 2 M NaCl and (C) 2 M urea.

3.3.8 Analysis of Samples Contaminated with SDS. SDS poses a particular concern for MALDI analysis. This substance is very difficult to remove by washing and the MALDI process can only tolerate very low amounts of SDS [15,35,36]. In this experiment, several protein solutions were prepared in 0.01% and 0.02% SDS. The samples were digested with trypsin both in free solution, and following the bead preconcentration procedure. Figures 3.6 and 3.7 summarize the results obtained for the digestion of 500 nM and 100 nM cytochrome c prepared in SDS. When Figure 3.6A and 3.6B are compared, one can see that the presence of 0.01% SDS in the sample has a serious effect on the resulting MALDI spectrum. Although the intensity of the obtained signals remained high in Figure 3.6B, the number of peaks and therefore the sequence coverage was greatly reduced. For 500 nM cytochrome c digested in solution without SDS present, 17 peaks were observed resulting in 68% coverage. When 0.01% SDS was added to the sample, only 7 peaks were detected, resulting in 29% sequence coverage. When the spectrum obtained from the digest of cytochrome c with 0.02% SDS in the sample was analyzed (Figure 3.6C), the quality of the spectrum was severely affected. A number of unidentifiable peaks are now present, and these peaks almost completely mask the few cytochrome c peaks that remain. If this were an unknown sample, the resulting spectrum would be indecipherable. One can therefore see that the presence of even low amounts of SDS has a serious effect on the quality of the spectrum.

Recall from Chapter 2, section 2.3.9, where protein samples were digested and the peptide fragments were analyzed in the presence of 0.1% SDS, no severe loss in signal intensity was observed. In that case, the protein samples were prepared at a concentration of 1 μg/μl. In this case, the concentration of protein was much lower (500 nM

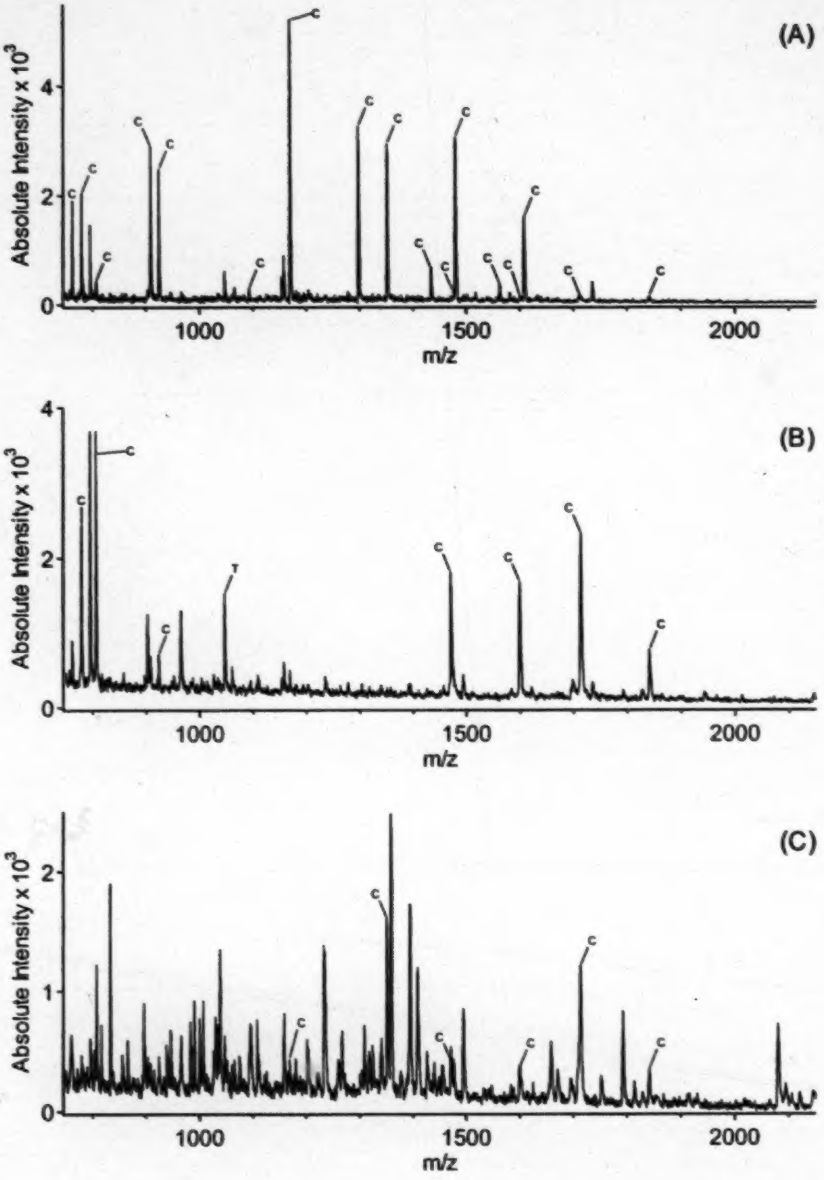

Figure 3.6 MALDI spectra obtained from the free solution digestion of 500-nM cytochrome c in (A) pure water, (B) 0.01% SDS and (C) 0.02% SDS.

cytochrome c or 0.006 μg/μL). Thus, for the diluted protein solution, the relative ratio of SDS to protein is actually much higher. Also, the digestion efficiency of a 500 nM protein solution is significantly reduced, as compared to digesting a higher concentration of protein, and then diluting to lower levels. Thus, the actual concentration of peptides that are detected by MALDI MS is much lower than 500 nM. As mentioned in section 3.1, contaminants become an increasing concern as we approach the detection limit of the approach [1]. These reasons combine to explain the detrimental effect of even low concentrations of SDS for peptide mass mapping of dilute protein samples.

The MALDI spectra obtained from the bead digests of 100 nM cytochrome c are displayed in Figure 3.7. Figure 3.7A is for the digest of cytochrome c in the absence of SDS. A total of 29 peaks are observed, representing 100% sequence coverage. Figure 3.7B shows that with 0.01% SDS in the sample, essentially all the cytochrome c peptide fragments observed from the digestion of the clean sample are still obtained. At 0.02% SDS (Figure 3.7C) the signal intensity dropped considerably, yet most of the more intense peaks in the spectrum are identified as cytochrome c fragments, with 24 of the 29 peaks originally detected from the clean sample being observed. However, when the concentration of SDS is higher than ~0.03%, no useful spectra of the trypsin digest of 100 nM protein solutions can be obtained with the method. These results demonstrate that the bead washing and digestion procedure can tolerate a modest amount of SDS (up to 0.02%). More significantly, however, is that the method will tolerate a high relative amount of SDS to protein. As an example, a 1 μM cytochrome c solution prepared in 0.2% SDS revealed similar results to the 100-nM sample in 0.02% SDS. Samples with higher levels of SDS can simply be diluted in water prior to the bead-digestion procedure, provided that the concentration of protein is also sufficient.

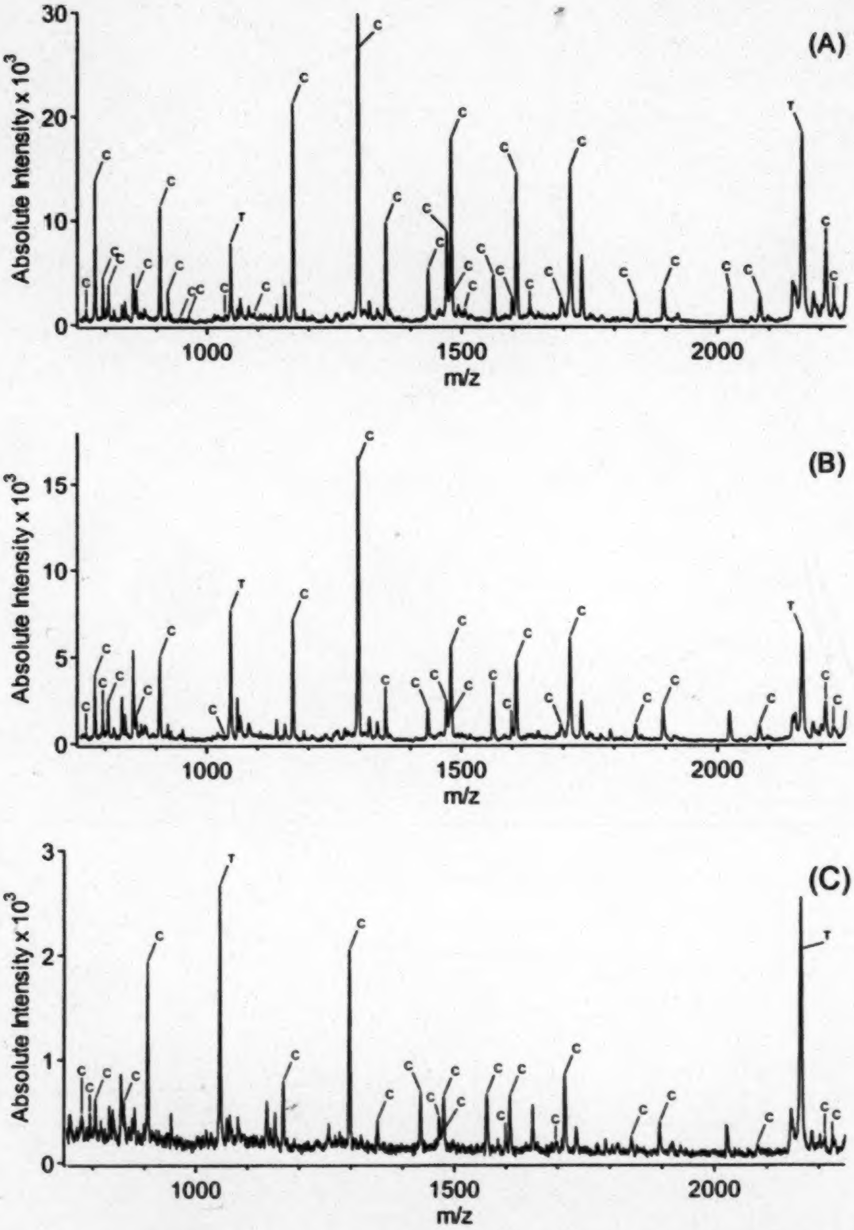

Figure 3.7 MALDI spectra obtained from micro-bead preconcentration and digestion of 100-nM cytochrome c in (A) pure water, (B) 0.01% SDS and (C) 0.02% SDS.

3.3.9 On-Probe Washing *vs.* Bead Washing. NaCl and urea are highly water soluble, and will not co-crystallize with the matrix and sample on the target. These contaminants will precipitate on top of the matrix-sample layer and therefore can be easily removed by on-probe washing with water. Other salts, such as NH_4HCO_3, behave in a similar fashion. In our experiments, we found that it was necessary to incorporate an on-probe washing step for the samples digested in solution, otherwise no signals corresponding to peptide fragments could be obtained. For spectra recorded without on-probe wash, only matrix clusters were observed. Therefore, this washing step was effective at removing contaminants that otherwise had a severe effect on the MALDI spectra. However, the on-probe washing step could not effectively remove the SDS and these spectra suffered as a result. It appears that SDS will associate with the peptides in the sample, and therefore washing with water becomes ineffective.

Using the bead digestion procedure, we are unable to perform an on-probe washing, as the beads would be removed in the process. However, we can make use of the difference in binding strength of the protein sample and various contaminants with the bead surface to wash the sample prior to loading the beads on the probe. The protein sample binds very strongly to the bead surface and therefore the beads can be washed extensively with little loss of sample. During the washing process, weakly bound substances are removed from the bead surface. We are therefore able to selectively concentrate dilute protein samples without concentrating the contaminants.

3.3.10 Resolution and Accuracy of MALDI Analysis. One of the key steps in the procedure is the loading of the beads on the MALDI probe. The method described in the experimental section was optimized to achieve maximal signal intensities and shot-to-shot reproducibility. We also found that forming a uniform bead layer on the probe can minimize the spatial effect on mass resolution and mass measurement accuracy in MALDI-TOF. In the linear time-lag focusing instrument, it was found that mass resolution was reduced by approximately 25%, when the beads were added to the probe. However, in a reflectron time-lag focusing instrument (*i.e.* Voyager DE Elite), baseline isotope resolution was easily achieved with the bead experiment, and the presence of beads on the sample plate did not appear to reduce the resolution.

It was also found that, using trypsin autolysis peaks to perform a two point internal calibration, the mass accuracy on the linear time-lag focusing instrument was 200 ± 100 ppm for sample preparations both in free solution and with beads. On the reflectron instrument with the same two point internal calibration, a mass accuracy of 40 ± 20 ppm was obtained for each preparation. From these results, we see that the mass accuracy is not affected by the bead digestion and detection procedure. An improvement in mass accuracy can be made by using several trypsin autolysis peaks to perform a multi-point calibration [37]. However, since it is often the case that no more than two autolysis peaks are seen in the spectra (the peaks at m/z ~2163 and 1046), the practice of using multiple points is not always possible.

3.4 Conclusions

Using a concentration procedure that involves adsorption of proteins onto polymeric micro-beads, peptide mass maps were successfully generated from dilute protein solutions. The advantage of this method over traditional preconcentration methods is that the procedure can also tolerate high concentrations of contaminants. The preconcentration and washing steps are rapid, and the increase in protein concentration prior to digestion also improves the kinetics of the enzyme digestion, so that the digestion of proteins from a 100 nM solution only takes 30 minutes to complete. The resulting peptide mass maps have broad sequence coverage and high signal intensities.

This method opens new opportunities to detect and characterize trace amount of proteins by mass spectrometry. The use of nano-scale technology has been explored to significantly reduce the number of beads required per sample. By reducing the starting volume required to concentrate a sample, a lower absolute amount detection limit is achieved. With this method, a 50-fmol sample of BSA (100 nL of a 500 nM solution) has been digested on beads.

3.5 Literature Cited

(1) Keller, B. O.; Li, L. *J. Am. Soc. Mass Spectrom.* **2000,** *11*, 88-93.

(2) Papac, D. I.; Hoyes, J.; Tomer, K. B. *Anal. Chem.* **1994,** *66*, 2609-2613.

(3) Zhou, W.; Merrick, B. A.; Khaledi M. G.; Tomer, K. B. *J. Am. Soc. Mass Spectrom.* **2000,** *11*, 273-282.

(4) Zhao, Y.; Chait, B. T. *Anal. Chem.* **1994,** *66*, 3723-3726.

(5) Nelson, R. W.; Krone, J. R.; Bieber, A. L.; Williams, P. *Anal. Chem.* **1995,** *67*, 1153-1158.

(6) Schriemer, D. C.; Li, L. *Anal. Chem.* **1996,** *68*, 3382-3387.

(7) Schriemer, D. C.; Yalcin, T.; Li, L. *Anal. Chem.* **1998,** *70*, 1569-1575.

(8) Girault, S.; Chassaing, G.; Blais, J. C.; Brunot, A.; Bolbach, G. *Anal. Chem.* **1996,** *68*, 2122-2126.

(9) Hurst, G. B.; Buchanan, M. V.; Foote, L. J.; Kennel, S. J. *Anal. Chem.* **1999,** *71*, 4727-4733.

(10) Hutchens, T. W.; Yip, T. T. *Rapid Commun. Mass Spectrom.* **1993,** *7*, 576-580.

(11) Brockman, A. H.; Orlando, R. *Anal. Chem.* **1995**, *67*, 4581-4585.

(12) Liang, X.; Lubman, D. M.; Rossi, D. T.; Nordblom, G. D.; Barksdale, C. M. *Anal. Chem.* **1998**, *70*, 498-503.

(13) Bundy, J.; Fenselau, C. *Anal. Chem.* **1999**, *71*, 1460-1463.

(14) Gevaert, K.; Demol, H.; Puype, M.; Broekaert, D.; De Boeck, S.; Houthaeve, T.; Vandekerckhove, J. *Electrophoresis* **1997**, *18*, 2950-2960.

(15) Gevaert, K.; Demol, H.; Sklyarova, T.; Vandekerckhove, J.; Houthaeve, T. *Electrophoresis* **1998**, *19*, 909-917.

(16) Aguilar, M.; Clayton, D. J.; Holt, P.; Kronina, V.; Boysen, R. I.; Purcell, A. W.; Hearn, M. T. W. *Anal. Chem.* **1998**, *70*, 5010-5018.

(17) Amankwa, L. N.; Kuhr, W. G. *Anal. Chem.* **1992**, *64*, 1610-1613.

(18) Amankwa, L. N.; Kuhr, W. G. *Anal. Chem.* **1993**, *65*, 2693-2697.

(19) Licklider, L.; Kuhr, W. G. *Anal. Chem.* **1994**, *66*, 4400-4407.

(20) Licklider, L.; Kuhr, W. G.; Lacey, M. P.; Keough, T.; Purdon, M. P.; Takigiku, R. *Anal. Chem.* **1995**, *67*, 4170-4177.

(21) Hsieh, Y. L. F.; Wang, H.; Elicone, C.; Mark, J.; Martin, S. A.; Regnier, F. *Anal. Chem.* **1996**, *68*, 455-462.

(22) Blackburn, R. K.; Anderegg, R. J. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 483-494.

(23) Gobom, J.; Nordhoff, E.; Ekman, R.; Roepstorff, P. *Int. J. Mass Spectrom. Ion Processes.* **1997**, *169*, 153-163.

(24) Ekstrom, S.; Onnerfjord, P.; Nilsson, J.; Bengtsson, M.; Laurell, T.; Marko-Varga, G. *Anal. Chem.* **2000**, *72*, 286-293.

(25) Keller, B.; Li, L. unpublished results on PerSeptive Biosystems' immobilized trypsin **1999**.

(26) Dai, Y.; Whittal, R. M.; Li, L. *Anal. Chem.* **1999**, *71*, 1087 –1091

(27) Whittal, R. M.; Li, L. *Anal. Chem.* **1995**, *76*, 1950-1954.

(28) Walker, A. K.; Land, C. M.; Kinsel, G. R.; Nelson, K. D. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 62-68.

(29) Lu, X.M.; Figueroa, A.; Karger, B. L. *J. Am. Chem. Soc.* **1998**, *110*, 1978-1979.

(30) Jilge, G.; Janzen, R.; Giesche, H.; Unger, K. K.; Kinkel, J. N.; Hearn, M. T. W. *J. Chromatogr.* **1987**, *397*, 71-80.

(31) Benedek, K.; Dong, S.; Karger, B. L. *J. Chromatogr.* **1984**, *317*, 227-243.

(32) Sane, S. U., Cramer, S. M., Przybycien, T. M. *J. Chromatogr., A* **1999**, *849*, 149-159.

(33) McNay, J. L.; Fernandez, E. J. *J. Chromatogr., A* **1999**, *849*, 135-148.

(34) Purcell, A. W.; Aguilar, M.; Hearn, M. T. W. *Anal. Chem.* **1999**, *71*, 2240-2451.

(35) Jenson, C.; Haebel, S.; Andersen, S. O.; Roepstorff, P. *Int. J. Mass Spectrom. Ion Processes* **1997**, *160*, 339-356.

(36) Jeannot, M. A.; Zheng, J.; Li, L. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 512-520.

(37) Whittal, R. M.; Russon, L. M.; Weinberger, S. R.; Li, L. *Anal. Chem.* **1997**, *69*, 2147-2153.

Chapter 4

Investigation of the Effects of Surface-Type on the Digestion of Surface Bound Proteins in Micro-Columns [5, 6]

4.1 Introduction

In Chapter 3, it was demonstrated that protein samples could be efficiently digested with trypsin following adsorption of the protein to a polymeric micro-bead support [1,2]. This technique was shown to be particularly useful for the purpose of generating MALDI peptide mass maps from dilute protein samples. Although the results shown in Chapter 3 clearly illustrate that enzymatic degradation of the bound protein has occurred, it is expected that the process of protein adsorption to a surface will alter the degradation process. To this point, little is known regarding how protein adsorption can affect its enzymatic degradation. For an adsorbed protein, interactions between the protein and the hydrophobic surface may lead to inaccessibility of the enzyme to cleave some, or even all possible digestion sites of the protein. This would lead to incomplete degradation of the protein, which may potentially result in a peptide mass map having low sequence coverage. However, for proteome analysis, high sequence coverage is generally desirable. The confidence of protein identification using the peptide mass

---

[5] A form of this chapter has been submitted for publication as: A. Doucette, D Craft, and L. Li "Effects of Hydrophobic Micro-bead Type on the Tryptic Digestion of Surface-Bound Proteins".

[6] D. Craft made significant contributions towards the development and optimization of the in-column digestion procedure described in this chapter. He also collected the electrospray results used in Figures 4.4-4.6.

mapping approach generally increases as the extent of amino acid sequence coverage of a protein by the digested peptides increases. In addition, since post-translational modifications (PTM's) of proteins are very common, high sequence coverage is also very important in identifying possible PTM's and determining the modified site(s) of a protein.

In some related work, proteins bound to antibodies have been digested with enzymes as a means of mapping epitopes. It has been shown that the rate of enzymatic cleavage of a protein complex is reduced, most significantly at regions where the protein is in contact with the antibody (the epitope) [3,4]. Similarly our previous results on biotinylated proteins complexed to avidin-coated beads revealed that such interactions prevented efficient trypsin digestion [5,6]. However, these interactions are quite different from those that are experienced by a protein adsorbed to a hydrophobic surface.

In a recent report, Aguilar *et al.* proposed that for a protein adsorbed to reversed-phase bonded silica, only those sites exposed to the solvent would be accessible to proteolysis [7]. They reported that the trypsin digestion of cytochrome c adsorbed to reversed-phase C_{18} beads was blocked at certain amino acid residues that correspond to regions of protein contact with the surface. In our previous work described in Chapter 3, we compared a digest on beads with a solution phase digestion. All peptides detected by MALDI from the solution digest were also detected in the bead digestion, indicating that the adsorption process did not prevent cleavage of the protein. However, the relative rate of digestion for each cleavage site may still be altered by the adsorption process; we note that the relative quantity of the peptides detected in the previous experiments were unknown [1]. The different conclusions drawn from the two reports may simply reflect differences in the experimental conditions, such as the type of beads used and the method of detection for analyzing the resulting digest.

In this chapter, we present our studies on the effects of experimental variables, particularly the type of micro beads, on the digestion efficiency of surface-bound proteins. The digestion products are analyzed by direct MALDI alone, or a combination of liquid chromatography/electrospray ionization mass spectrometry (LC/ESI-MS), direct MALDI, LC/offline MALDI-MS, and UV absorption. In addition, a new method is presented for the digestion of surface bound proteins. The method involves the capture of proteins from dilute solutions into micro-columns packed with various types of chromatographic supports. This in-column digestion procedure offers improved sample manipulation as compared to the bead digestion procedure described in Chapter 3, and has the potential to be fully automated. A detailed examination of the in-column digestion procedure is presented here.

4.2 Experimental

4.2.1 Materials and Reagents. Bovine serum albumin (BSA), equine cytochrome c, and trypsin (from bovine pancreas, TPCK treated to reduce chymotrypsin activity; dialyzed, lyophilized) were from Sigma Aldrich Canada (Oakville, ON). Poros 20 R2 beads (20 μm) were from PerSeptive Biosystems (a gift of Professor F. Cantwell, University of Alberta). PRP-3 (12–20 μm), PRP-1 (12–20 μm), and PRP-Infinity (4 μm) beads were purchased from Hamilton Company (Reno, NV). Reversed-phase C_{18}, C_8, and C_4 bonded-phase silica beads (20-30 μm) were from Vydac (Hesperia, Ca). The 1-200 μL Micro capillary Tips (gel loader tips) used to form the columns were purchased from Rose Scientific (Edmonton, AB). The glass micro-fiber filters used to form column frits were purchased from Whatman (Maidstone, Eng.). Analytical grade acetone, methanol, acetonitrile, acetic acid, and trifluoroacetic acid (TFA) were purchased from Caledon Laboratories (Edmonton, AB). Water used in the experiments was from a NANOpure water system (Barnstead/Thermolyne). All other chemicals were purchased from Sigma Aldrich Canada.

4.2.2 Protein Digestion in Micro-columns. The design of the column is based on work presented by Roepstorff *et al.* [8] and by Annan *et al.* [9]. Briefly, an aliquot of the packing material (chromatography beads), suspended in methanol at a concentration of 10 mg/mL, was transferred to a Micro capillary gel loader tip (1-200 μL), which had been pinched at the end with pliers. A 1×1-mm piece of glass micro-fiber membrane, placed at the end of the micro-capillary tip, was used as a frit to ensure that the micro-beads did not pass through the tips. The tip was then briefly centrifuged to form a packed column. The procedure for sample loading and digestion is based on our previous work [2]. Briefly, a 100-μL protein sample, at a concentration of 100 nM, was pumped through the column a total of three times at an approximate flow rate of 10 μL/min, thus allowing 30 minutes for sample loading. The column was then washed with 100 μL of water. For the BSA sample, a 1 μM BSA solution was first treated with DTT and iodoacetamide to break disulfide bonds according to standard protocols. The reduced BSA was then diluted in water to 100 nM for loading on the column. The absorbed protein was digested by flowing a 10-μL solution of 31-ng/μL trypsin in 50 mM NH_4HCO_3 through the column. A small volume of enzyme solution was left in the column to prevent the beads from drying and aid in digestion. The column was incubated for 30 minutes at 37°C and, following digestion, the sample was eluted into a vial using 3 μL of 50% acetonitrile/ 0.1% TFA. Each column was used once and then discarded.

4.2.3 Digestion of Protein Adsorbed on Micro-Beads. Poros R2 beads and Vydac C_{18} silica beads were saturated with cytochrome c and subjected to proteolytic cleavage with trypsin using a procedure similar to that described by Aguilar *et al.* [7]. The beads (7.5 mg) were first washed in methanol and then in 35% acetonitrile and suspended for 2 hours in 0.5 mL of solution containing 7.5-mg/mL cytochrome c. Following this, the beads were centrifuged and then washed four times in enzyme buffer (50 mM NH_4HCO_3 containing 2 mM $CaCl_2$). The adsorbed cytochrome c was then subjected to trypsin digestion by incubating the beads in 0.5 mL of enzyme buffer containing 75 μg of trypsin for 20 hours at 37°C with shaking. The digestion was terminated with 2 drops of 2 M HCl and the supernatant was retained. The beads were then eluted with 250 μL each of 25%, 50%, and 75% acetonitrile in 0.1% TFA, pooling all the eluates for analysis. 1-μL portions of the supernatant and of the extract were subjected to MALDI analysis. The supernatant and eluate from a given digest were then pooled and concentrated by solvent evaporation with a Savant SpeedVac concentrator to a final volume of approximately 200 μL. The sample was then subjected to HPLC separation and analysis.

4.2.4 Solution Digestion of Protein. For the purpose of comparison to the bead digests, a solution-phase digestion of cytochrome c was performed. Cytochrome c was prepared at a concentration of 1 μg/μL in water. To 10 μL of the protein sample, 1 μL of 1 M NH_4HCO_3 and 1 μL of 1-μg/μL trypsin were added. The solution was briefly vortexed and then incubated at 37°C for 2 hours. The digestion was stopped by acidifying with 0.5 μL of 10% TFA.

4.2.5 HPLC Separations and Fractionation. Peptides were separated by HPLC with a Hewlett Packard Series 1100 chromatographic system (Agilent Technologies, Palo Alto, Ca). The column was a Vydac reversed-phase n-octylsilica (C_8) column (Hesperia, CA) with dimensions of 250 × 4.6-mm i.d.. The sorbent particle size was 5 μm with 300-Å pores. A gradient elution was performed using 0.02% (v/v) TFA in water (buffer A) and 0.02% (v/v) TFA in acetonitrile (buffer B). For online mass spectrometric detection with an ion trap, 0.5% acetic acid was used in the solvents in place of TFA. The gradient used was as follows: hold B at 2% for 5 minutes, then increase from 2 to 50% B over the next 48 minutes, all at a constant flow rate of 1 mL/min. For the cytochrome c digested on beads, a given run consisted of approximately one quarter of the extracted sample injected on the column. For the solution phase digest, 10 μL of the sample (8.3 μg cytochrome c) was injected on the column. UV chromatograms were recorded at 210 nm. For offline analysis of the chromatographic separation by MALDI, the eluent was collected in vials at one-minute intervals with an automated fraction collector.

4.2.6 MALDI Sample Preparation. The matrix α-cyano-4-hydroxycinnamic acid (HCCA) was used in these experiments. HCCA was first purified by recrystallization from ethanol. Two sample preparation techniques were employed to deposit the samples for MALDI analysis. For the column digests, a 0.4 μL portion of undissolved HCCA suspended in 50% acetonitrile/0.1% TFA was added to the 3 μL extract of the column digest, thus ensuring saturation of the sample with matrix. The sample-matrix suspension was vortexed briefly and then centrifuged for 2 minutes to settle any undissolved matrix. A 0.7 μL portion of the solution was deposited on a MALDI target and allowed to dry in air. No on-probe washing was performed on the dried sample spot. For the HPLC fractions, the samples were concentrated by solvent evaporation with a Savant Speed Vac concentrator (Fisher Scientific) from 1 mL down to approximately 50 μL. The evaporated samples were then analyzed by MALDI using a two-layer deposition method [10]. The sample was mixed with second layer matrix in a ratio of 1:2. Following sample deposition and drying, on-probe washing was performed by placing 1 μL of water on the MALDI sample spot, and blowing off the water with air after a few seconds.

4.2.7 Mass Spectrometry. MALDI spectra of the in-column digests were collected on a linear time-lag focusing MALDI time-of-flight mass spectrometer, which was constructed at the University of Alberta and has been described in detail elsewhere [11]. The fractions from the HPLC separation were recorded using a Bruker Reflex III MALDI time-of-flight system (Bremen/Leipzig, Germany). With each instrument, a pulsed nitrogen laser operating at 337 nm was used to generate the MALDI ions. All spectra shown in this chapter are the results of signal averaging of between 100 and 200 shots. Data processing was performed with the IGOR Pro software package (Wavemetrics Inc., Lake Oswego, OR).

On-line liquid chromatography/electrospray ionization mass spectrometry (LC/ESI-MS analysis was performed on a Finnigan LCQ^{deca} ion trap mass spectrometer

(Finnigan, San Jose, CA). The effluent from the column was split, allowing approximately 20 μL/min to flow to the ion trap.

4.3 Results and Discussion

4.3.1 In-Column Digestion. Packing the beads in a column offers several advantages over the addition of beads directly to the sample vial. In addition to avoiding sample loss due to inefficient transfer of beads, samples can be easily washed in the column, thereby reducing weakly bound contaminants prior to digestion. The simple washing step incorporated here involves flowing 100 μL of water through the column prior to digestion. This method offers a more convenient approach to washing compared with our previous work, where the beads were placed in a sample vial, pelleted by centrifugation, and the supernatant removed without disturbing the settled beads. Also, for the column digestion procedure, an on-probe washing step can be incorporated following sample deposition on the MALDI target to further eliminate contaminants, particularly those salts associated with the HCCA matrix solution itself. In the previous method, the deposition of beads directly on the MALDI target prevented any form of on-probe washing.

Figure 4.1 illustrates the performance on the in-column digestion procedure for solutions of 100 nM cytochrome c, prepared in pure water, 2 M NaCl, or 2 M urea. The solutions were eluted through columns packed with R2 beads. Prior to digestion, the column was washed with 100 μL of water to remove any weakly bound contaminants. From Figure 4.1A, we can see that the application of the in-column digestion procedure results in the detection of 30 peptide fragments of cytochrome c in the MALDI spectrum, representing a total of 100% sequence coverage. The number and intensity of the

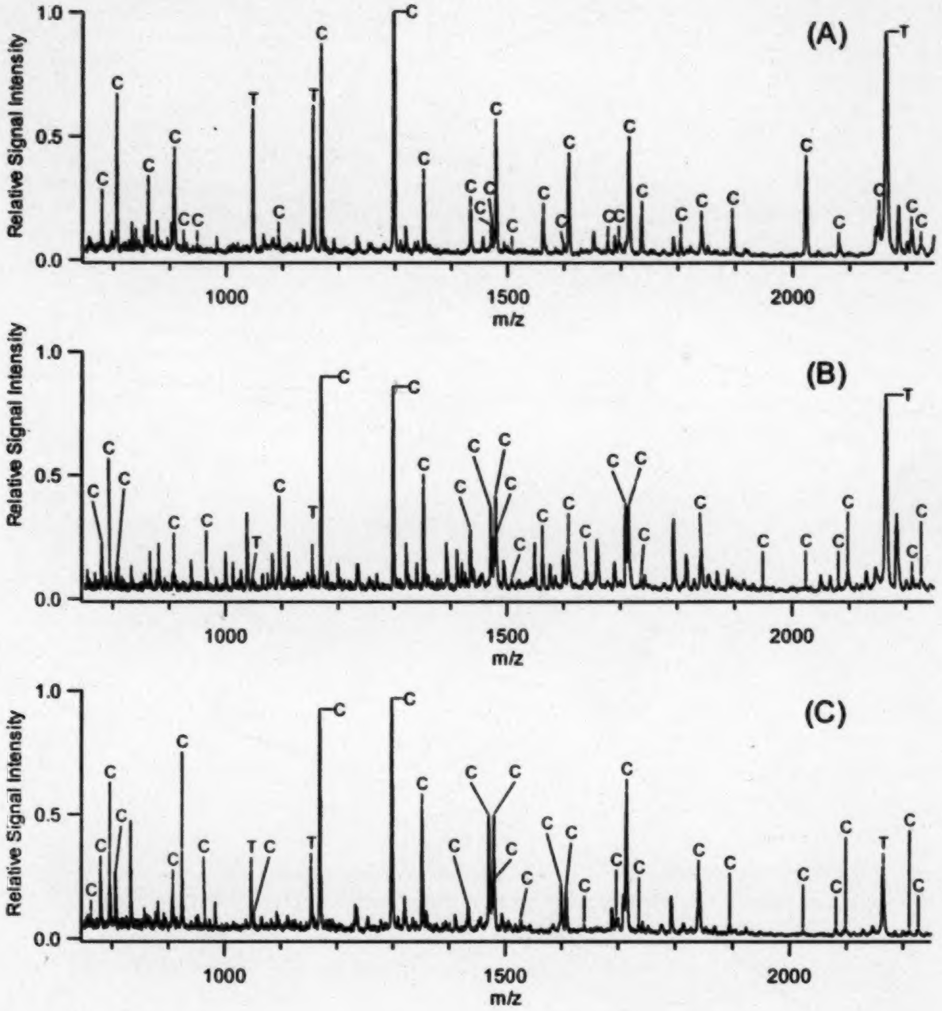

Figure 4.1 MALDI spectra showing the products of the digestion of 100 nM cytochrome c in-column that was packed with R2 beads for a (A) clean solution, (B) solution containing 2 M urea, 29 peaks were detected, giving 100% coverage, (C) solution containing 2 M NaCl, 27 peaks were detected, giving 97% coverage, and (D) with 0.01% SDS, 27 peaks were detected, giving 97% coverage. C = peptide fragments of cytochrome c, T = peptide fragments resulting from trypsin autolysis.

detected peaks are similar to that obtained using the bead digestion procedure described in Chapter 3 (see Figure 3.5).

In addition, the samples containing high concentrations of contaminants (NaCl or urea) also display high quality MALDI spectra using the in-column digestion procedure to pre-concentrate and clean the protein samples. For the sample containing NaCl (Figure 4.1B), a total of 29 peptides were assigned to fragments of cytochrome c representing 100% sequence coverage. The sample contaminated with urea displayed 27 peaks (Figure 4.1C), representing 97% sequence coverage. The quality of the peptide mass map for the salt- and urea-containing samples is essentially constant when compared to the uncontaminated cytochrome c digest of Figure 4.1A. These contaminants are largely unretained during the sample-loading phase, and are further removed from the column by incorporation of a washing step. As a result, no significant degradation in signal is obtained during the MALDI analysis. Samples of 100-nM BSA or lysozyme were also prepared in 2 M NaCl and 2 M urea. The MALDI spectra obtained from the analysis of these digests were essentially identical to those obtained from the corresponding "clean" protein standards.

These results demonstrate that the in-column digestion procedure can be effectively employed for the digestion and subsequent detection of dilute protein samples. The in-column digestion protocol can thus be used to study the effects of surface type on the digestion of adsorbed proteins.

4.3.2 Detailed Examination of Surface Effects on Digestion Efficiency. The enzymatic degradation of proteins adsorbed on a surface may be affected in several different ways. It is possible that certain sites of a surface-bound protein may not be cleaved by the enzyme, which would result in an increased number of incompletely digested peptides (*i.e.*, larger fragments of the protein). On the other hand, surface bonding could influence the protein folding in such a manner that would result in increased accessibility of the enzyme to certain cleavage sites. In this case, the resulting digest should contain many peptides covering a wide range of cleavage sites. In addition to altering the relative digestion rate of certain cleavage sites due to surface bonding, the overall digestion efficiency of the entire surface-bound protein may also be altered. This can be attributed to steric hindrance arising from the necessity of an enzyme to change its conformer as it approaches the surface and interacts with the surface-bound protein. We have designed several experiments to examine these effects and the results are presented below.

4.3.3 In-Column Digestion of Proteins on Various Supports. Figure 4.2 shows the MALDI spectra obtained from the digests of cytochrome c adsorbed to various supports. In this experiment, micro-columns were prepared using the following four types of micro-beads: Poros R2 beads, C_{18} bonded-phase silica beads, C_8 beads, as well as C_4 beads.

As seen in Figure 4.2, many peaks corresponding to peptide fragments of cytochrome c are visible in the MALDI spectra (these peaks are labeled "C" in the figure). A number of the cytochrome c fragments were observed in each of the four surface digests, although they display significant differences in relative signal intensities in the MALDI spectra. Along with the peptide fragments of cytochrome c, other peaks are also visible, including trypsin autolysis fragments (labeled "T" in the figure) as well as a few commonly observed matrix clusters of HCCA. In addition, several unidentified peaks (the unlabeled peaks in the figure) are also visible in the spectra displayed in Figure 4.2. In particular, these unknown peaks increase in number and intensity as the support

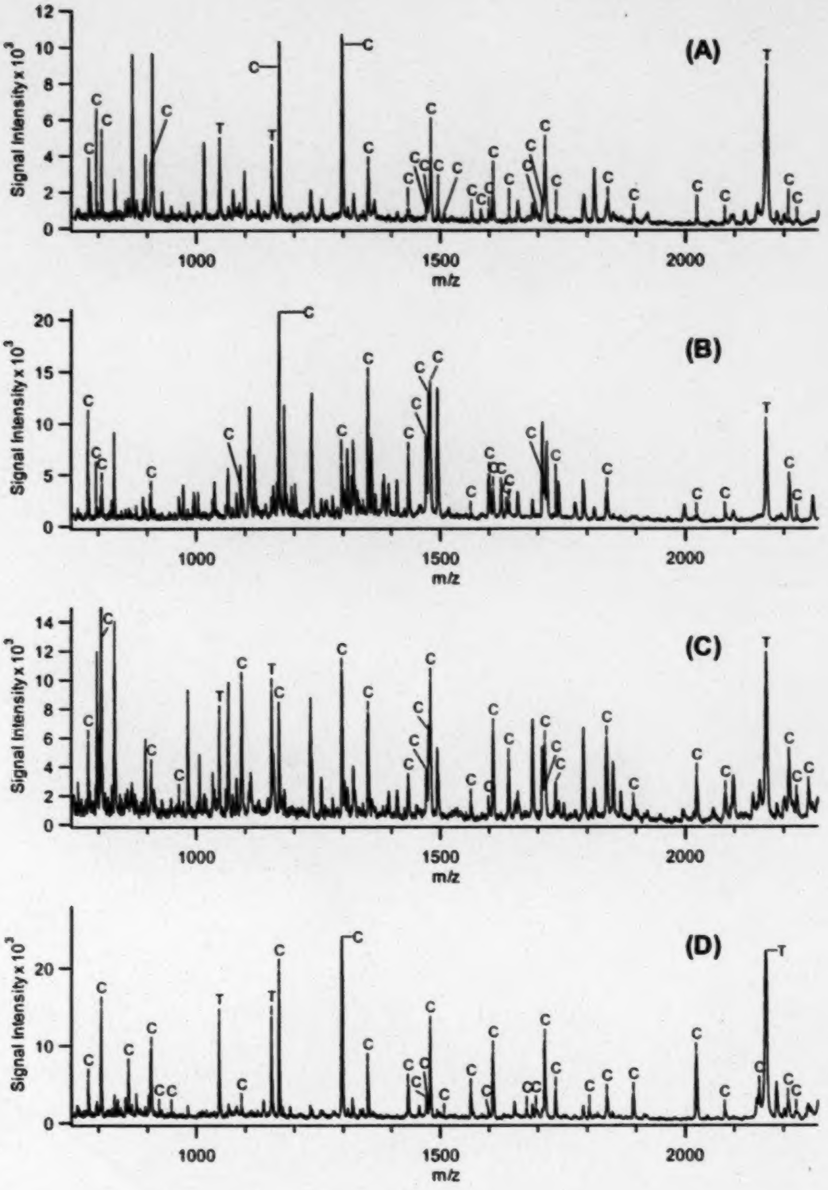

Figure 4.2 MALDI spectra showing the in-column digestion of 100 nM cytochrome c, where the column was packed with (A) Vydac C_{18}, (B) Vydac C_8, (C) Vydac C_4, and (D) Poros R2 beads. In the spectra, the label "C" refers to peptide fragments of cytochrome c, and "T" refers to peptide fragments resulting from trypsin autolysis.

used in the digestion procedure is varied from R2 to C_{18} to C_8 to C_4. It is suspected that these peaks may reflect fragments of cytochrome c, arising from non-specific tryptic cleavage of the protein while absorbed to the various supports. These unidentified peaks, along with other factors such as the variation in signal intensities, contribute to noticeable differences in the MALDI spectra for the various bead supports.

Although the spectra shown in Figure 4.2 exhibit several differences, the information conveyed by the MALDI spectra for each surface digest is very similar. This is more clearly shown by the data presented in Table 4.1, which lists the peptide fragments detected from each of the four column digestions. From this table, we see that the obtained peptide mass maps from the cytochrome c digests on each type of support are very similar. For each digest, a similar number of peptide fragments were detected in the MALDI spectrum, ranging from 22 to 28 fragments. Combining these peptides, a pooled total of 33 unique fragments were detected. Of the unique fragments, 19 were detected in each of the MALDI spectra. These 19 common peptides correspond to 88.5% sequence coverage of the protein (missing the coverage of amino acid residues 14 through 25). Of the peptide fragments that were not commonly observed in all four spectra, most were of low signal intensity.

Besides cytochrome c, other proteins can also be effectively digested when bound to various hydrophobic surfaces. Table 4.2 shows the detected peptides from column digests using the same four types of supports from a representative example, bovine serum albumin (BSA). As shown in Table 4.2, several peptide fragments were detected from the four column digestions. Between 32 and 36 peaks were detected in each individual digest, resulting in a pooled total of 46 unique peptides. In this case, 26 fragments were common to the four spectra. The 26 common peptides correspond to

Table 4.1 Tryptic peptides detected from the digests of cytochrome c bound to various surfaces.

peak #	amino acid residue	peak mass	Type of hydrophobic surface			
			C_{18}	C_8	C_4	R2
1	1-8	860.5				X
2	1-13	1475.9	X	X	X	X
3	6-13	947.6				X
4	9-22	2250.8			X	
5	14-22	1635.4		X		
6	23-38	1675.9				X
7	23-39	1804				X
8	26-39	1561.9	X	X	X	X
9	26-38	1433.8	X	X	X	X
10	28-38	1168.6	X	X	X	X
11	28-39	1296.7	X	X	X	X
12	39-53	1598.8	X	X	X	X
13	39-55	1840.9	X	X	X	X
14	40-53	1470.7	X	X	X	X
15	40-55	1712.8	X	X	X	X
16	56-72	2081	X	X	X	X
17	56-73	2209.3	X	X	X	X
18	61-72	1495.7	X			
19	61-73	1623.8		X		
20	73-79	806.5	X	X	X	X
21	74-88	1695	X			X
22	80-86	779.4	X	X	X	X
23	80-87	907.5	X	X	X	X
24	87-100	1735	X	X	X	X
25	87-104	2150.2				X
26	88-99/ 89-100	1478.8	X	X	X	X
27	87-99/ 88-100	1606.9	X	X	X	X
28	88-104	2022.1	X	X	X	X
29	89-99	1350.7	X	X	X	X
30	89-104	1894	X		X	X
31	92-99	964.5			X	
32	92-100	1092.6		X	X	X
33	92-104	1507.8	X			X
total number of peaks			23 peaks	22 peaks	23 peaks	28 peaks
sequence coverage			89%	97%	97%	91%

Table 4.2 Tryptic peptides detected from the digests of BSA bound to various surfaces.

peak #	amino acid residue	peak mass	Type of hydrophobic surface			
			C-18	C-8	C-4	R2
1	25-34	1193.6	x	x	x	x
2	29-34	712.4	x	x	x	
3	35-44	1249.6	x		x	x
4	45-88	4910.4			x	
5	66-75	1163.6	x	x	x	x
6	101-105	545.3		x	x	
7	161-167	927.5	x	x	x	x
8	161-168	1083.6	x	x	x	x
9	168-183	2045.0	x	x	x	x
10	198-218	2289.2				x
11	205-211	906.5	x	x	x	x
12	205-222/ 210-228	2144.2		x	x	x
13	219-222	572.4	x	x	x	
14	229-232	508.3	x	x	x	x
15	219-235	2004.1	x	x	x	x
16	223-232	1138.6	x	x	x	x
17	233-241	1001.6	x	x	x	x
18	236-241	689.4	x	x	x	x
19	242-248	847.5	x	x	x	x
20	249-266	2057.2	x			
21	257-266	1153.7	x	x	x	x
22	257-285	3211.6	x		x	x
23	300-309	1177.6		x		
24	347-359	1567.7	x	x	x	x
25	347-360	1723.8	x	x	x	x
26	347-374	3300.8	x	x	x	x
27	360-371	1438.8	x	x	x	x
28	361-371	1283.7	x		x	x
29	400-420	2539.3	x	x	x	x
30	400-433	4000.1	x	x	x	x
31	402-412	1305.7	x	x	x	x
32	421-433	1479.8	x	x	x	x
33	421-436	1900.0	x	x	x	x
34	437-451	1639.9	x	x	x	x
35	437-455	2025.2		x		
36	437-459	2438.4		x		x
37	438-451	1511.8		x		x
38	452-459	817.5		x		
39	452-482	3515.7				x
40	548-557	1142.7	x	x	x	x
41	549-557	1014.6	x	x	x	x
42	549-561	1504.9		x		
43	549-587	4407.2			x	x
44	558-568	1308.7	x	x	x	x
45	569-580	1399.7				x
46	569-597	3153.4			x	
total number of peaks			32 peaks	36 peaks	36 peaks	36 peaks
sequence coverage			38%	36%	47%	48%

sequence coverage for BSA of 31%. As seen in Table 4.2, additional peptide fragments from the individual digests increased the total sequence coverage to between 36% and 48%. As for cytochrome c, the unique BSA peaks detected in a single support type were only weakly observed in the MALDI spectra.

The results from Tables 4.1 and 4.2 indicate that the peptide mass map obtained by direct MALDI analysis of the digest gives a similar level of sequence coverage from the four surfaces examined and no single surface produces a superior digest. For protein identification by peptide mass mapping, any one of the surfaces can be used for column digestion. However, some unique peptide peaks were observed from each surface, although their signals in the direct MALDI spectra were generally weak. Since MALDI results are not strictly quantitative, the intensities of these weak peaks do not reflect the actual peptide amounts in the digest. Furthermore, direct MALDI analysis of a complex peptide mixture may not detect all components of the mixture; a result of both signal suppression, as well as the concentration of certain fragments falling below their detection limit. It is clear that the use of an appropriate peptide analysis technique is important to reveal the inherent differences in digestion processes of proteins bound to various surfaces. We chose to use a combination of several detection techniques with a hope that these different modes of detection could reveal any bias towards the detection of peptides by a particular technique. These techniques include LC/UV, LC/ESI-MS, direct MALDI-MS, and LC fractionation followed by offline MALDI-MS.

4.3.4 HPLC Separation with UV detection of Peptide Fragments. In order to provide a more quantitative analysis of the digestion products from various surfaces, HPLC separation with UV absorbance detection of the digestion products of cytochrome c was performed. In this experiment, cytochrome c was adsorbed and digested on R2 or

C_{18} beads, using a procedure similar to that described by Aguilar [7]. For comparison, a solution-phase digestion of cytochrome c was also performed. The resulting chromatograms are presented in Figure 4.3.

Figure 4.3A displays the UV chromatogram for the peptide fragments of the solution digest of cytochrome c. Several well-resolved peaks of varying intensities are visible in this trace. However, many peaks, such as the peak at 20.3 minutes, are only partially resolved, revealing that the separation did not resolve all components of the mixture (MS analysis of this separation confirmed that a single fraction often contained several peptide fragments). The UV chromatograms for the surface-bound digestions on C_{18} and R2 beads are presented in Figure 4.3B and 4.3C respectively. When they are compared to the UV chromatogram from the solution digest, it is apparent that protein adsorption to a surface must alter the tryptic digestion of cytochrome c. Figure 4.3 shows that, although several peaks of similar intensities are observed in both the solution digest and the surface digests (such as the peaks near 10, 20, 21, and 29 min), other peaks have significant variations in signal intensities between the UV chromatograms. In particular, the chromatograms from the surface digests (Figure 4.3B and 4.3C) show several intense peaks between 35 and 40 minutes, which are almost absent in the chromatogram of the solution digest. As well, from Figure 4.3A, the solution digest displays various peaks earlier in the separation (near 16, 18, and 22 min) whose intensities are significantly reduced or are absent in the UV chromatograms of the surface digests. The late-eluting peaks represent more hydrophobic, often larger peptide fragments, while those peptides eluting early in the separation are more hydrophilic, often shorter peptide fragments. The results from Figure 4.3 therefore indicate that tryptic digestion of a protein adsorbed to a

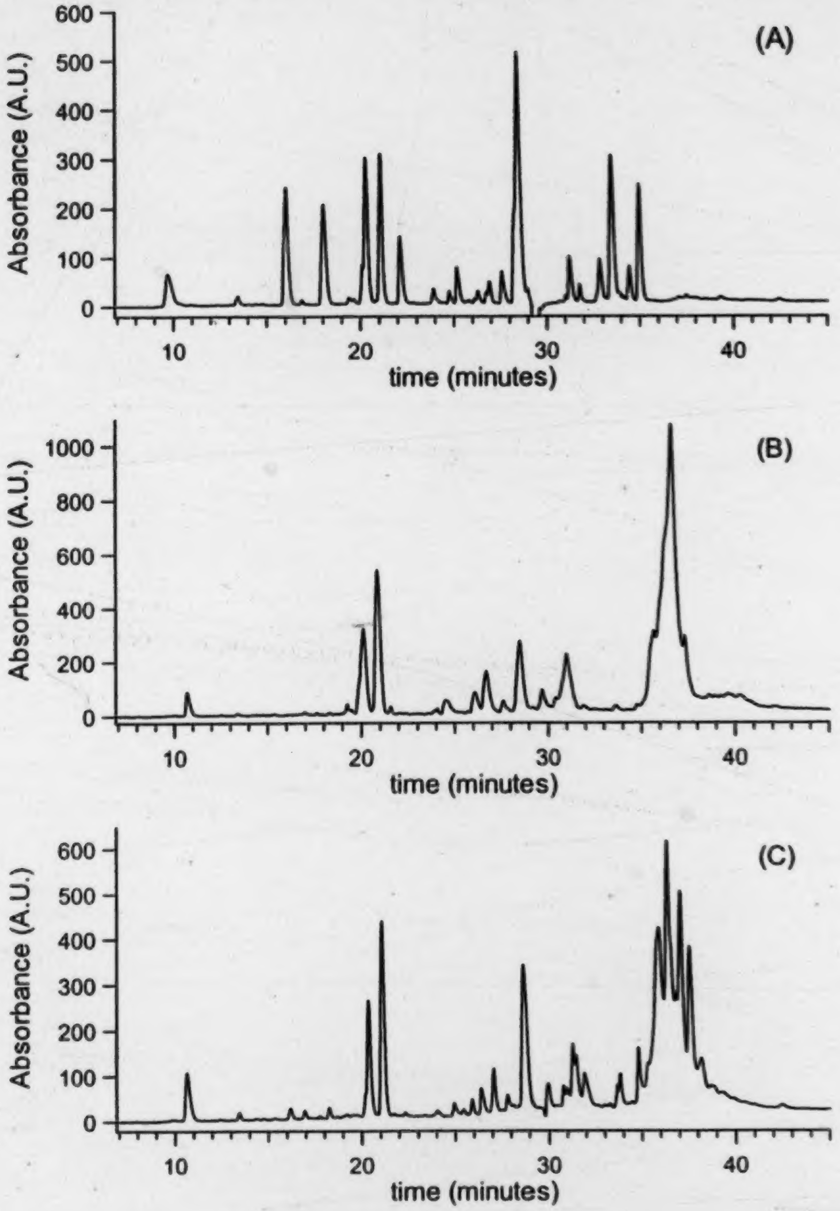

Figure 4.3 UV chromatograms for the HPLC separations of peptide fragments from the digestion of cytochrome c: (A) in solution, (B) on C_{18} silica beads, and (C) on polymeric R2 beads.

surface produces an increased proportion of larger, more incompletely digested peptide fragments.

Although the UV chromatograms of the solution and surface-bound digests display clear variations, the chromatograms from the digestion of cytochrome c on C_{18} and on R2 beads (Figure 4.3B and 4.3C) strongly resemble each other. More notably, the relative UV absorbencies of the peaks are similar. This therefore demonstrates that the relative abundance of many of the peptide fragments in each digest are similar. Some differences are observed from the UV chromatograms for the digests of the surface-bound samples. Figure 4.3C reveals a slightly higher relative amount of early eluting peptides in the UV chromatogram of the R2 digest, as compared to the C_{18} digest of Figure 4.3B. Also, the relative intensities of the partly resolved peaks between 35 and 40 minutes in the UV chromatogram are noticeably different from those of the C_{18} digest. As it was determined by MS analysis that individual peaks in the UV chromatogram are often composed of several peptides, it becomes difficult to relate these changes in signal intensity to specific changes in the relative amounts of individual peptides. However, the results of the UV chromatograms do indicate that the type of surface has some influence on the digestion of surface bound proteins. More specifically, it appears that an increased proportion of smaller peptide fragments are generated from the digestion of cytochrome c on R2 beads, as compared to C_{18} beads. We next performed MS analysis of the digestion products under identical separation conditions as those used in the generation of the UV chromatograms.

4.3.5 MS Detection of Solution and Surface-Bound Protein Digests. Three modes of MS detection were carried out for analyzing the protein digests. Cytochrome c, in solution or adsorbed to R2 or C_{18} beads, was first digested with trypsin. Following the bead digestions, the aqueous supernatant, as well as the extraction solvents were individually analyzed by direct MALDI. The extract and supernatant was then combined and approximately one fourth of the total sample was separated by HPLC. The HPLC fractions were collected and individual fractions were analyzed by MALDI (*i.e.*, LC/offline MALDI-MS). Equal portions of the samples were also analyzed online by LC/ESI-MS using similar separation conditions. All peptide fragments of cytochrome c observed under these three detection methods were recorded and the results are summarized in Figures 4.4-4.6.

Figure 4.4 displays the peptides observed from the solution digest of cytochrome c. Figure 4.5 contains the peptide fragments from the C_{18} digestion of cytochrome c, while Figure 4.6 contains the data for the R2 digests of the same protein. The peptide fragments from the MALDI and ESI experiments are displayed separately in each of these figures. Also, those peptides that were uniquely detected in either the direct MALDI analysis or the LC/offline MALDI analysis are specified in the figures. The first column of each figure lists the tryptic cleavage sites in the cytochrome c sequence. A total of 22 fragments are expected from a complete tryptic digest of cytochrome c. However, partial cleavage of the protein, which results in the generation of peptide fragments containing missed cleavages, could give rise to a greater number of fragments. The following columns of each figure show the peptides detected by LC/offline MALDI and direct MALDI, as well as by LC/ESI-MS, with the listed numbers corresponding to the number of missed cleavages.

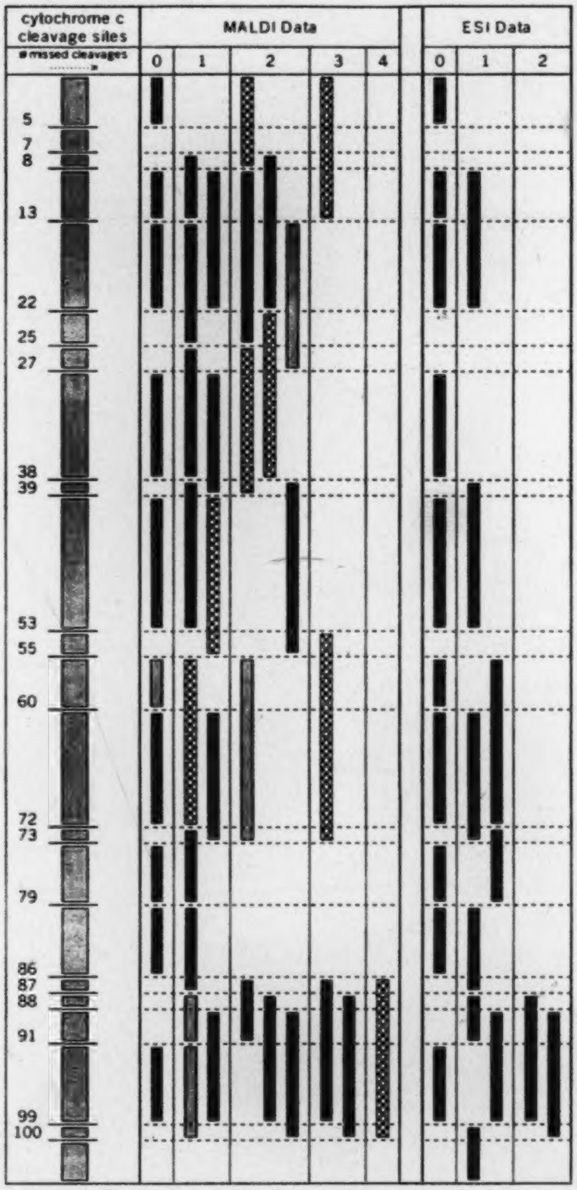

Figure 4.4 Tryptic peptides detected from the digest of cytochrome c in solution by using MALDI and ESI-MS. The fragments observed exclusively in the direct MALDI analysis are indicated with a light gray bar, while those exclusive to LC/offline MALDI are indicated with a dark gray bar. Fragments commonly observed in each MALDI analysis method are indicated with a black bar. The number of missed cleavage sites per peptide are indicated in the figure.

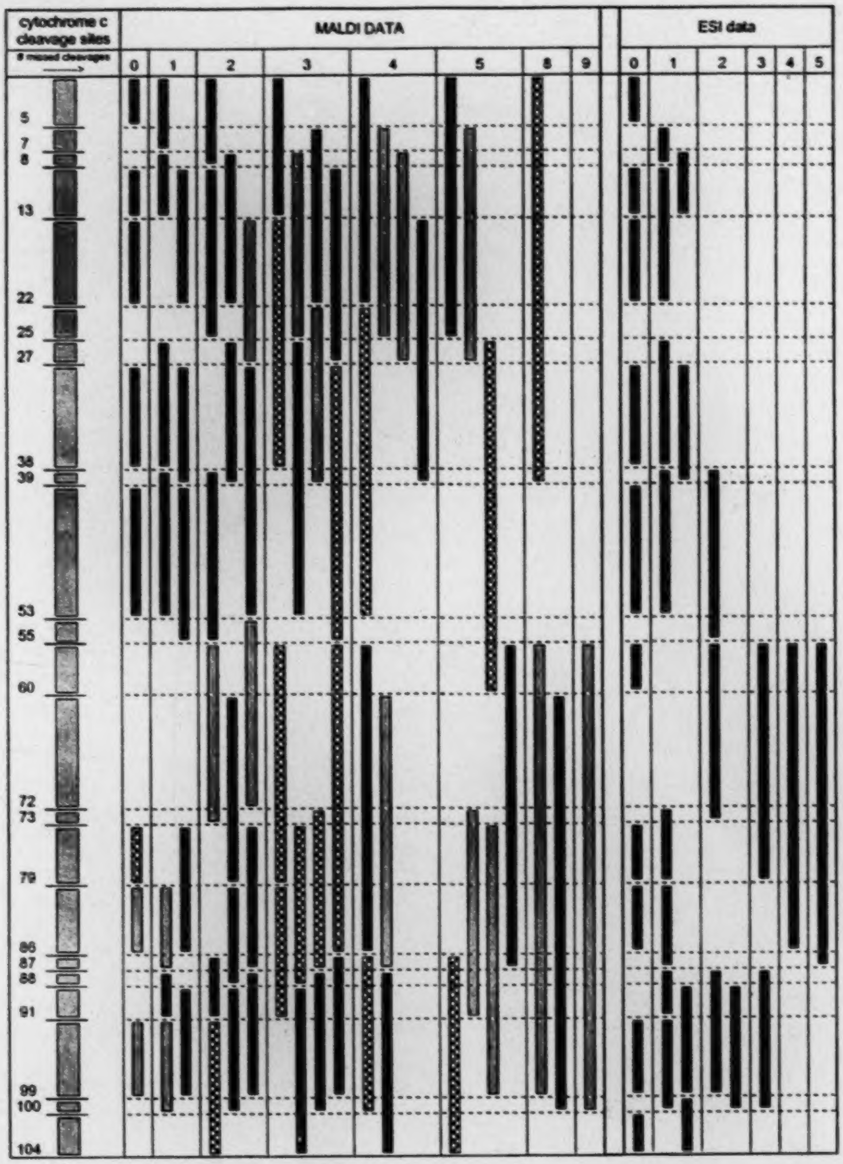

Figure 4.5 Tryptic peptides detected from the digest of cytochrome c bound to the C_{18} surface by using MALDI and ESI-MS. The fragments observed exclusively in the direct MALDI analysis are indicated with a light gray bar, while those exclusive to LC/offline MALDI are indicated with a dark gray bar. Fragments commonly observed in each MALDI analysis method are indicated with a black bar. The number of missed cleavage sites per peptide are indicated in the figure.

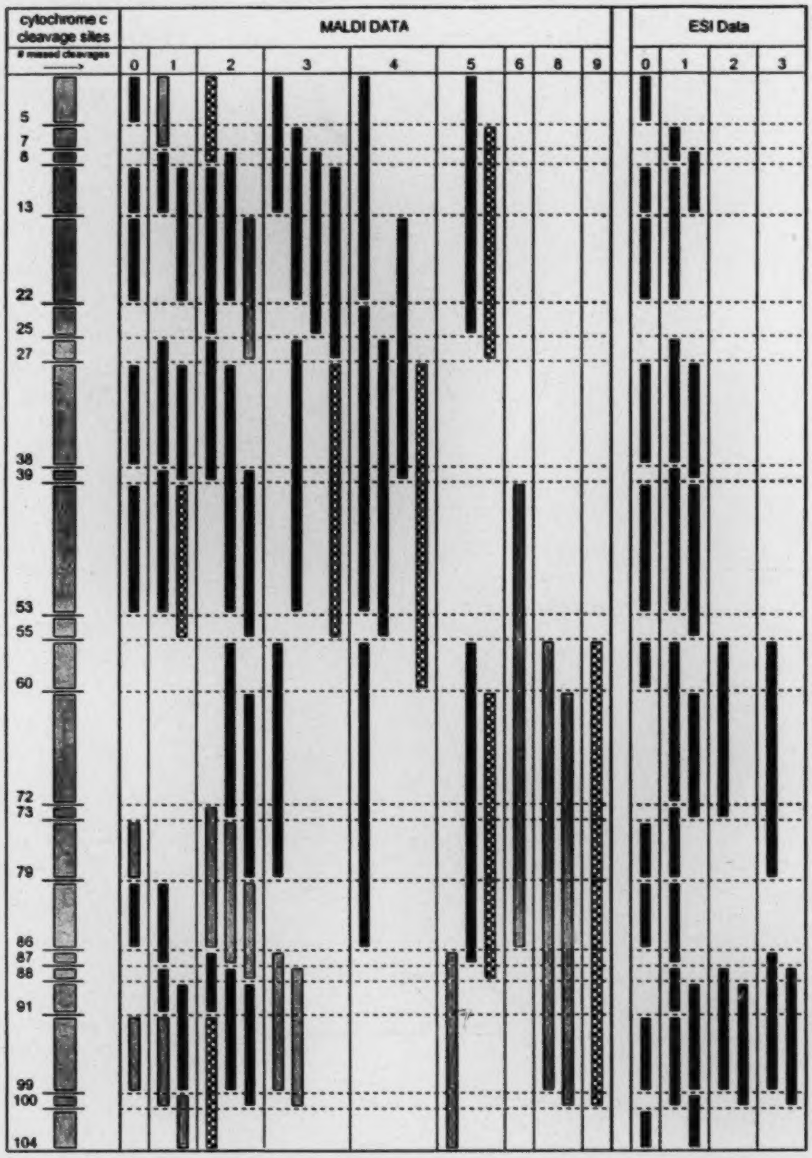

Figure 4.6 Tryptic peptides detected from the digest of cytochrome c bound to the R2 surface by using MALDI and ESI-MS. The fragments observed exclusively in the direct MALDI analysis are indicated with a light gray bar, while those exclusive to LC/offline MALDI are indicated with a dark bar. Fragments commonly observed in each MALDI analysis method are indicated with a black bar. The number of missed cleavage sites per peptide are indicated in the figure.

As observed in Figure 4.4, under the conditions used, a solution digestion of cytochrome c results in a partial digest of the protein, since many of the detected peptides contain one or more missed cleavage sites. This figure also shows that the greatest number of missed cleavages observed on a single peptide fragment is 4, for a peptide spanning amino acid residues 87 to 100. The partial digestion of cytochrome c results in a high number of peptides being generated. From Figure 4.4, 32 peptides were detected by direct MALDI analysis of the solution digest, 35 peptides were detected from LC/offline MALDI, while 21 peptides were detected by LC/ESI-MS. From the combination of the three detection schemes, a total of 41 unique peptides were detected.

The data displayed in Figure 4.4 also provides information on the ability of the various MS methods to detect peptides. First, it is noted that for each MS analysis method used, only a subset of the total number of peptides present in the digest were detected. Of the 41 unique fragments, only 17 were common to all three detection schemes. Also, for each detection method, at least one peptide was observed exclusively in that particular detection method. Therefore, the most complete analysis of the peptides in a digest would include all three methods of analysis. It is likely that additional peptide fragments present in the sample mixture remained undetected by all MS methods. In particular, short peptide fragments (spanning from 2 to 4 amino acid residues), are difficult to detect by MALDI or ESI MS owing to their low molecular weight, or inability to be retained on a reversed-phase separation column.

A comparison of the peptides detected from the various MS methods reveals a bias between the detection methods. Figures 4.4 through 4.6 show that MALDI detection reveals a larger number of peptides than LC/ESI-MS. This is attributed to an increased number of higher molecular weight peptides (generally those having increased missed cleavages) detected by MALDI-MS but not with LC/ESI-MS. Related to this, those peptides that were exclusively detected by LC/ESI-MS were mainly low molecular weight peptides that are difficult to observe in a MALDI spectrum. Within the two MALDI detection schemes, again only certain peptides were detected with each method. It was expected that LC/offline MALDI-MS would increase the total number of observed peptide fragments, due to a decrease in the ion suppression effect. However, several peptides observed in the direct MALDI analysis were no longer observed following sample fractionation. A possible explanation is related to the fact that for the direct MALDI detection, the supernatant and the eluate of the bead digests were analyzed separately. Also, LC fractionation did not resolve all components of the digest mixture. As a result, the co-elution of a subset of peptides in which one or more component is strongly ionized may still suppress the low abundant, or weakly ionizable peptides. Nonetheless, the combined MS methods can successfully detect a greater number of peptide fragments, hence providing a better overall picture of the peptide mass map of the digest.

Combining the results of the three detection schemes, 100% sequence coverage of the protein was obtained. Also, from the peptide fragments detected, one can observe that all possible digestion sites of cytochrome c were cleaved, at least to some degree, by trypsin. Limited cleavage of individual digestion sites can be inferred from the data of Figure 4.4. As seen in this figure, peptide fragments containing at least one missed cleavage site are observed for all lysine and arginine residues of cytochrome c. The results of Figure 4.4 show that, under the conditions used, partial protein degradation is observed even in a solution digest of cytochrome c with trypsin.

The results of the solution digest of cytochrome c can be compared to the peptide fragments detected from the digestion of cytochrome c adsorbed on the supports. Figures 4.5 and 4.6 display the peptide fragments detected from the C_{18} and R2 bead digests of cytochrome c. It is immediately obvious that a tryptic digest of cytochrome c on these supports results in the generation of a larger number of detectable peptide fragments, as compared to the solution digest of the protein. Combining both the ESI and MALDI data, a total of 66 unique peptide fragments were detected from the R2 bead digest (Figure 4.6), while 77 unique peptide fragments were observed from a C_{18} bead digest of cytochrome c (Figure 4.5). This high number of peptide fragments results from a partial cleavage of each individual tryptic cleavage site of cytochrome c. The degree of cleavage of each site is observed to be lower than the solution digestion of cytochrome c, since much larger peptide fragments, having as many as 8 missed cleavage sites on a single fragment, were also observed from the bead digests of cytochrome c. Most of the larger fragments observed in the bead digests, containing at least three missed cleavages, were not observed from the solution digest of the sample. In addition, some of the peptides containing no, or only one missed cleavage were not present in the bead digest. However, these peptides were detected from a solution digestion of cytochrome c. In particular, the results of Figures 4.5 and 4.6 reveal a gap in the amino acid sequence detected from the MALDI analysis of the more completely digested peptides between residues 54 to 73. ESI results reveal only a portion of this sequence, which can be attributed to increased detection sensitivity of these low mass peptides.

The bead digests of Figure 4.5 and 4.6 display an increased proportion of peptide fragments, having a higher degree of missed cleavage sites than those of the solution digest of cytochrome c. This can be attributed to a decrease in the degree of digestion of individual cleavage sites of the protein. This conclusion is deduced from reasoning as follows: the probability of generating a peptide fragment containing a high number of missed cleavages will increase as the degree of digestion of each cleavage site decreases. In related reasoning, the probability of generating peptides with little or no missed cleavages increases as the degree of digestion increases. Thus, for the bead digests in which the smaller fragments covering residues 54 to 73 are not detected, whereas larger fragments are detected covering this same sequence of the protein, it is concluded that the degree of cleavage of these digestion sites is significantly reduced as compared to a solution phase digestion of the protein. Thus, the adsorption of cytochrome c to R2 or C_{18} beads can result in reduced digestion efficiency, which is more pronounced at certain regions of the cytochrome c. This is likely due to the proximity of contact between regions of the protein and the surface of the solid support.

It should be noted that although the digestion efficiency decreases as the cytochrome c is adsorbed on a solid support, all cleavage sites of cytochrome c were digested to some degree while bound to the R2 or C_{18} supports. This is demonstrated in Figures 4.5 and 4.6 where peptide fragments were detected with terminating residues for all 21 lysine and arginine residues of cytochrome c. Therefore, the adsorption of cytochrome c to a solid support does not completely block the digestion of any individual digestion site of the protein. Also, 100% sequence coverage was obtained from the digests on each type of support. As a related observation, all amino acid residues of the protein were covered by a minimum of two unique peptides detected by either LC/MALDI or LC/ESI-MS. The most extreme representation of a single segment of the protein was for residue 14 to 22, which was detected from 16 unique peptide fragments in the C_{18} bead digest of cytochrome c. It is interesting to note that this segment of amino acid residues was completely missed from the C_{18} column digest of 100 nM cytochrome c, presented in Table 4.1. These results therefore demonstrate that both the concentration of peptide in the sample, as well as the detection method employed (fractionation prior to MALDI analysis or MALDI analysis of the raw mixture) will greatly influence the ability to detect a given peptide in the sample.

From the comparison of the peptide fragments in Figures 4.4 – 4.6, we can see that the digestion of cytochrome c adsorbed to a support will produce a peptide map containing a wider range of peptide fragments as compared to a solution digestion of the protein. This result is significant in that additional sequence coverage can often be obtained by the generation of different populations of peptide fragments for a given protein. Just as various enzymes can be used to generate unique peptide fragments for a protein, the adsorption of a protein to a surface will generate unique peptides from a digestion with the same enzyme. Peptide fragments containing a higher degree of missed cleavage sites are useful to confirm the sequence of a segment of protein observed from smaller peptide fragments. In other cases, segments of a given protein can sometimes only be observed by MALDI or ESI through these larger fragments with increased missed cleavages. Where cleavage sites are closely spaced within a given protein, a complete digestion of the molecule would create short segments, often containing less than 4 amino acid residues. These small fragments are often not detected by MALDI or ESI, due to their low molecular weight or extremes in hydrophobicity. This is particularly useful where the detection of all residues is required, as in the case of amino acid sequencing or the localization of post-translational modifications (PTM's). The applications of this approach for characterization of PTM's will be reported elsewhere.

The effect of surface type on the digestion of absorbed protein can be explored by comparing the data presented in Figure 4.5 and 4.6. As mentioned in earlier discussion, an increased number of peptides containing missed cleavages can be attributed to a decrease in the digestion efficiency of individual cleavage sites of the protein. From Figure 4.5, a total of 77 unique fragments were detected from a C_{18} bead digest whereas 66 unique peptides were detected from the R2 bead digest (Figure 4.6). Also, a larger proportion of peptide fragments containing a greater number of missed cleavage sites was observed from the C_{18} digest as compared to the R2 bead digest. Of the 66 unique peptides detected in the R2 extract, 24 peptides (36%) contained 3 or more missed cleavages. This compares to 47% (36 out of 77 peptides) having 3 or more missed cleavages from the C_{18} digest. These results indicate that the digestion efficiency of cytochrome c while absorbed to C_{18} beads is slightly lower compared to the digestion efficiency on R2 beads. However, the effect of support type on the digestion efficiency is relatively small compared to the differences resulting between a solution digestion and a digest of a surface-bound protein. From Figure 4.4, only 12% of the peptides detected (5 out of 41 unique fragments) contained 3 or more missed cleavage sites.

4.3.6 Effects of Bead Construction and Pore Size on Digestion. In addition to possible effects of surface type on digestion efficiency, it is possible that for a chemically similar support, variations in bead construction may lead to differences in the digests. In particular, some protein might enter pores in the beads, making the protein inaccessible to trypsin. To address this, we compared the peptide maps obtained from the digestion of cytochrome c on various polystyrene divinylbenzene surfaces, each having differences in the structure of the bead, namely in pore size. The beads used in this study were from Hamilton Company, including PRP-3 beads (mean pore size 300 Å, 12–20 μm diameter), PRP-1 beads (mean pore size 100 Å, 12 – 20 μm diameter), and PRP-Infinity beads (non-porous, 4 μm diameter), as well as Perseptive Biosystems Poros R2 beads (having through-pores of 6000-8000 Å, and diffusive pores of 800-1500 Å, bead diameter 20 μm). The obtained MALDI spectra from digests using these beads were all similar (data not shown). In particular, samples digested on PRP-Infinity beads did not appear to improve the quality of the MALDI spectrum, compared to samples digested on beads having varying pore sizes. Therefore, it is concluded that pore size does not have a significant effect on the digestion efficiency of an adsorbed protein.

4.4 Conclusions

Protein samples can be digested with trypsin while adsorbed on hydrophobic surfaces, namely reversed-phase chromatography beads. Peptide mass maps obtained by direct MALDI analysis show similar levels of protein sequence coverage for the digests of proteins bound to C_{18}, C_8, C_4, and R2 beads. In addition, the pore size of the beads does not have a significant effect on the digestion efficiency of an adsorbed protein. Thus, for protein identification by peptide mass mapping using direct MALDI analysis of the surface-bound protein digest, any one of the hydrophobic surfaces examined herein can be used. The choice of the micro-bead surface is therefore mainly dependent on the efficiency of protein retention, which can be optimized for different applications. For example, C_{18} beads should provide a better retention for many proteins and thus they are the preferred media for routine in-column digestion experiments. However, subtle differences in peptide mass maps are observed from proteins bound to R2 and C_{18} beads. We demonstrate that using a combination of direct MALDI, LC/offline MALDI, and LC/ESI MS analysis, instead of using a single MS detection technique alone, better reveals these changes.

4.5 Literature Cited

(1) Doucette, A.; Craft, D.; Li, L. *Anal. Chem.* **2000**, *72*, 3355-3362.

(2) Craft, D.; Doucette, A.; Li, L. *Proceedings 48th ASMS Conference on Mass Spectrom. Allied Topics*, Long Beach CA, **2000**, 1429.

(3) Papac, D. I.; Hoyes, J.; Tomer, K. B. *Anal. Chem.* **1994**, *66*, 2609-2613.

(4) Zhao, Y.; Chait, B. T. *Anal. Chem.* **1994**, *66*, 3723-3726.

(5) Schreimer, D. C.; Li, L. *Anal. Chem.* **1996**, *68*, 3382-3387.

(6) Schreimer, D. C.; Yalcin, T.; Li, L. *Anal. Chem.* **1998**, *70*, 1569-1575.

(7) Aguilar, M.; Clayton, D. J.; Holt, P.; Kronina, V.; Boysen, R. I.; Purcell, A. W.; Hearn, M. T. W. *Anal. Chem.* **1998**, *70*, 5010-5018.

(8) Gobom, J.; Nordhoff, E.; Ekman, R.; Roepstorff, P. *Int. J. Mass Spectrom. Ion. Processes* **1997**, *169/170*, 153-163.

(9) Annan, R.S.; McNulty, D.E.; Carr, S.A. *Proceedings 44th ASMS Conference on Mass Spectrom. Allied Topics*, Portland OR, **1996**, 702.

(10) Dai, Y.; Whittal, R. M.; Li, L. *Anal. Chem.* **1999**, *71*, 1087-1091.

(11) Whittal, R. M.; Li, L. *Anal. Chem.* **1995**, 67, 1950-1954.

Chapter 5

Investigation of the Applicability of a Sequential Digestion Protocol Using Trypsin and Leucine Aminopeptidase M for Protein Identification by MALDI-TOF Mass Spectrometry [7, 8]

5.1 Introduction

The previous chapters described general procedures employed in the digestion and subsequent mass spectrometric analysis of protein samples for peptide mass fingerprinting. As described in these chapters, the generation of peptide fragments is a critical parameter in establishing the validity of the protein identification method. A desirable protocol would produce peptide maps with broad amino acid sequence coverage of a protein in order to provide the highest level of information available for database searching, as well as to deduce information on post-translational modifications.

Besides peptide mass fingerprinting, one can also make use of partial amino acid sequencing from the N- or C-terminus of the intact protein as a second dimension in

---

[7] A portion of this chapter has been published as: A. Doucette, and L. Li "Investigation of the Applicability of a Sequential Digestion Protocol Using Trypsin and Leucine Aminopeptidase M for Protein Identification by MALDI-TOF Mass Spectrometry" joint issues of *Eur. J. Mass Spectrom.* **2001**, *23*, 4424-345, and *Proteomics*, **2001**, *4*, 324-245.

[8] J. Stupak collected the data related to the results presented in Figure 5.10.

identifying an unknown protein. Amino acid sequencing of an intact protein could be accomplished either enzymatically [1,2], chemically [3-6], or via fragmentation techniques, such as post-source decay, linear delayed-extraction, or by MS/MS [7-9].

Another method to increase the information available for protein identification is to incorporate partial amino acid sequencing of the peptide fragments resulting from the initial digestion (peptide mass tags). Amino acid sequence information has been deduced using a variety of different methods. For example, the peptides can be partially sequenced by Edman degradation [10,11], using in-source decay following protein cleavage with cyanogen bromide [12], or by using tandem MS following trypsin digestion [13-14]. Tandem MS instruments based on MALDI ionization have become increasingly available, owing to the commercialization of new MS instruments [15,16].

The use of enzymes for peptide sequencing provides a simple procedure that takes advantage of the specificity of the enzyme for controlled degradation of the peptide. In this method, following tryptic digestion of the protein to form peptide fragments, an exopeptidase is used for successive cleavage of amino acids to generate peptide mass tags. Most of the work reported on these lines involves separation of peptide fragments from a tryptic digest prior to sequencing. For example, Sillard *et al.* used carboxypeptidases to sequence the C-terminal end of HPLC separated tryptic fragments [17]. However, we reported that, with optimization of the digestion conditions, it is possible to generate peptide ladders after exopeptidases are added to the crude tryptic digest of a protein [18]. We applied this method for deducing sequencing information in the study of protein modifications [18]. Korostensky *et al.* also applied exopeptidases to unseparated protein digests [19]. They reported an algorithm for protein identification using the mass tags generated from the sequential enzyme digestions. Siuzdak *et al.*

reported a similar experiment involving sequencing of trypsin-digested proteins with carboxypeptidase Y [20]. In both these works, the digestion was performed by adding the exopeptidase enzyme to the sample after deposition on the MALDI target [19,20].

In this work, we extend our previous work on sequential enzyme digestions in solution [18] by investigating the general applicability of the rapid procedure involving trypsin and leucine aminopeptidase M for protein identification. The goal of this systematic investigation is to provide us with a level of understanding of the attributes and limitations of the procedure. We demonstrate that, while the procedure is experimentally more involved, it can generally increase our confidence of protein identification based on trypsin-digestion mass fingerprinting using matrix-assisted laser desorption ionization mass spectrometry. The method is also applied for improving the sequence information that can be deduced from MS/MS fragmentation information in a MALDI-TOF quadrupole-time-of-flight (Q-TOF) MS instrument.

5.2 Experimental

5.2.1 Materials and Reagents. The protein samples, bovine serum albumin (BSA), equine cytochrome c, bovine α-lactalbumin, bovine lactoferrin, bovine β-lactoglobulin, chicken lysozyme, equine myoglobin, and bovine ubiquitin, as well as trypsin (from bovine pancreas, TPCK treated to reduce chymotrypsin activity; dialyzed, lyophilized) and leucine aminopeptidase M (LAP), EC number 3.4.11.2, (5.6 µg/µl slurry in 3.5 M $(NH_4)_2SO_4$, pH 7.7, with 10 mM $MgCl_2$), were obtained from Sigma Aldrich Canada (Oakville, ON). CD9 was prepared by Dr. Andrew Shaw's lab at the Cross Cancer Institute (Edmonton, AB) by lysing Raji/CD9 B-cell lymphocytes (2×10^8 cells) in 1%

NP-40 for 90 min at 4°C, followed by centrifugation at 14000g for 10 min. The lysate was then partially purified on an affinity column prepared with mAB 50H.19, a CD9 antibody obtained from a mouse hybridoma [21]. The bacterial sample (*Escherichia coli*, ATCC 9637) was from Dr. Randy Long of Edgewood RDE Center, Aberdeen Proving Ground, US. For peptide and protein extraction, a solvent suspension method was used [22]. Briefly, the lyophilized *E. coli* cells were suspended in 0.1% TFA to dissolve the proteins and the supernatant solution was fractionated by HPLC before enzymatic digestion.

Analytical grade acetone, methanol, acetonitrile, acetic acid, and trifluoroacetic acid (TFA) were purchased from Caledon Laboratories (Edmonton, AB). Water used in the experiment was from a NANOpure water system (Barnstead/Thermolyne). Polyacrylamide gels used for one-dimensional SDS PAGE were from three different sources; Pre-cast Ready Gels (16.5% Tris-Tricine) from Bio-Rad (Richmond, CA), Protein Mini-Gels (10-20% gradient) from Genomic Solutions Inc. (Chelmsford, MA), and 15% SDS-polyacrylamide gels of 1 mm thickness cast in house. Silver staining was performed using the silver staining kits provided by Genomic Solutions Inc. and Bio-Rad. All other chemicals were from Sigma Aldrich Canada and were used without further purification unless otherwise stated.

5.2.2 Enzyme Digests. Samples consisting of 1 μg protein in 11 μl of 0.1 M NH_4HCO_3 (0.09 μg/μL) were added to 1 μl of 45 mM D-Dithiothreitol (DTT), followed by incubation at 55°C for 15 minutes, which reduced the disulfide linkages. The cysteines were then carbamidoacetylated to prevent reformation of disulfide bonds using 1 μl of 100 mM iodoacetamide, which was allowed to react for 15 minutes at room temperature. After this, 0.2 μg of trypsin was added to the protein solutions and allowed to digest for 1 hour at 37°C. The total volume of the digestion solution was 15 µl. For the sensitivity study, samples of 0.1 µg protein were digested with 0.05 µg trypsin, following cysteine reduction. The digests were then further diluted with 0.1 M NH_4HCO_3 to the appropriate concentration before N-terminal cleavage. The *E. coli* fractions were digested for 30 minutes at 37°C with 0.05 µg trypsin, following cysteine reduction.

N-terminal digestion was then performed on the 1 µg samples of trypsin-digested protein. The LAP slurry was briefly vortexed before mixing 1 µl of the slurry (5.6 µg LAP) with 14 µl of the trypsin-digested protein solution. This solution was allowed to digest at room temperature for defined periods (usually 2, 10, and 30 minutes), after which a 1-µl portion of the solution was removed and added to 10 µl of saturated matrix solution, which terminated the LAP digestion. For the sensitivity study, as well as the *E. coli* fractions, the LAP slurry was diluted two-fold with water to a concentration of 2.8 µg/µl LAP. To these samples, 0.5 µl of the diluted LAP was added. This was again allowed to react for defined periods of time before sampling 1 µl of the digest into a saturated matrix solution (4.5 µl for sensitivity study, 2 µl for *E. coli* samples).

5.2.3 SDS-PAGE and In-Gel Digests. Protein samples were loaded onto SDS-polyacrylamide gels for one dimensional gel electrophoresis. 10-µl samples containing 1 µg protein were loaded into the sample wells. For the sensitivity study, the total protein loading on the gel was reduced to 100 ng. The proteins were stacked at 30.0 V for one hour, followed by gel separation using 90.0 V for 2.5 hours. Silver staining was performed using the procedure described with the Bio-Rad Silver Staining Plus Kit or the Genomic Solutions Inc. Minigel Silver Stain Kit.

After staining, the protein bands were excised from the gel and in-gel digested with trypsin using the procedure outlined as follows: The gel pieces containing the protein samples were first washed with several changes of 0.1 M NH_4HCO_3, establishing a pH near 8. The NH_4HCO_3 was then removed and the gel pieces were mashed in a sample vial using a smaller sample vial to squeeze the gel between the vials. The gel pieces were then rinsed from the bottom of the smaller sample vial with 50 μl of 10 mM DTT in 0.1 M NH_4HCO_3, and the mixture was allowed to react for 1 hour at 55°C. Most of the added liquid was then removed with a micropipette after which 50 μl of 55 mM iodoacetamide in 0.1 M NH_4HCO_3 was added to the gel pieces. After a 45-minute incubation period at room temperature, the liquid was again removed with a micropipette. To the gel pieces, 50 μl of 0.1 M NH_4HCO_3 and 1 μl of 1 μg/μl trypsin was added. This mixture was then allowed to digest for 1 hour at 37°C. For the sensitivity study, the digestion volume was reduced to approximately 25 μl of 0.1 M NH_4HCO_3 to which 0.15 μg trypsin was added. This digestion proceeded for 2 hours at 37°C.

The peptide fragments were extracted three times at room temperature using 50 μl of 50% acetonitrile/ 0.5% TFA, followed by a final extraction with 20 μl of acetonitrile (20 minutes with sonication per extraction). The extraction liquid was pooled and evaporated to dryness, and then reconstituted with 10 μl of 0.1 M NH_4HCO_3. N-terminal digestion was then performed on the 1 μg samples using 5.6 μg LAP, and using 0.4 μg LAP with the 100 ng samples.

5.2.4 MALDI Sample Preparation. The samples were prepared for MALDI analysis using a two-layer method developed in this laboratory [23]. The matrix used in these experiments was α-cyano-4-hydroxycinnamic acid (HCCA), which was first purified by recrystallization from ethanol. A microcrytalline matrix layer was first deposited on the probe tip via fast evaporation from a 1-µl solution of HCCA (12 mg/ml) dissolved in 80% acetone/ methanol. One microliter of the protein digest solution was mixed with 2 to 10 µl of a saturated HCCA solution in 40% methanol/ water with 0.5 % TFA. The protein-matrix solution was briefly centrifuged and a 0.5-µl portion was deposited on top of the first matrix layer. After the solvent had evaporated, the dried spot was washed three times with 1.5 µl of room temperature water.

5.2.5 Instrumentation. Mass spectra were collected on a time-lag focusing MALDI time-of-flight mass spectrometer, which was constructed at the University of Alberta and has been described in detail elsewhere [24]. A pulsed nitrogen laser operating at 337 nm was used to generate the MALDI ions. The reported mass spectra are the result of signal averaging of approximately 40 laser shots. Data were processed using the IGOR Pro software package (Wavemetrics Inc., Lake Oswego, OR). Post source decay experiments were done in the Applied Biosystems Voyager Elite MALDI time-of-flight mass spectrometer and the Bruker Reflex III MALDI TOF instrument. A Bruker electrospray ionization ion trap mass spectrometer equipped with a nanospray source was used for ESI MS and ESI MS/MS. MALDI MS/MS fragmentation experiments were performed on a MDS Sciex API Q-TOF MS instrument (Toronto, ON).

5.2.6 Interpretation of Spectra. The amino acid sequences of the protein samples used in this study were obtained from the SWISS-PROT database (http://www.expasy.ch). Theoretical tryptic digests of the protein samples were performed using the web-based program, PeptideMass, found at http://www.expasy.ch/sprot/peptide-mass.html. By way of subtracting the masses of the first, second, and third amino acids from all peptide fragments that matched the theoretical trypsin digest, theoretical N-terminal digests of the tryptic-peptides could be generated. MALDI spectra resulting from sequential trypsin and LAP digestion were overlaid, revealing those peaks that appeared in the latter spectra resulting from products of N-terminal cleavage. Peptide mass fingerprinting of the *E. coli* fractionated samples was performed using the program PeptIdent (http://www.expasy.ch/cgi-bin/peptident.pl). Theoretical fragments for MS/MS CID fragmentation was performed by entering the sequence of the fragmented peptide in the MS-Tag program available on the UCSF mass spectrometry group website at http://prospector.ucsf.edu/ucsfhtml4.0u/mstagfd.htm.

5.3 Results and Discussion

5.3.1 Sequential Digestion. The general performance of the sequential digestion procedure in terms of amino acid sequence information was studied using eight standard proteins. We chose these readily available proteins so that others, at the initial stage of implementing this protocol, can repeat the procedure presented in section 5.2 of this chapter and compare their resulting data with those shown here. These proteins represent a diverse range of structure and molecular weights ranging from approximately 8600 Da for ubiquitin to over 76 kDa for lactoferrin. Some of these proteins also possess disulfide linkages, which can complicate endoprotease digestion. All protein samples were treated with D-dithiothreitol and iodoacetamide to reduce and block possible disulfide linkages, and digested with trypsin using an identical procedure. The resulting MALDI mass spectra were compared to theoretical trypsin digests of the proteins in order to identify the resulting peaks. The peak list matching the theoretic digest is depicted in Table 5.1. A representative spectrum corresponding to the trypsin digestion of cytochrome c is depicted in Figure 5.1A.

Table 5.1 Observed peptide fragment masses resulting from the trypsin digestion of eight protein samples in solution, each at a concentration of 0.09 μg/μl.

ubiquitin	
717.43	1081.52
765.43	1523.78
1039.52	1789.35
1067.69	

cytochrome c	
762.47	1350.99
779.45	1433.68
806.56	1470.69
907.94	1478.79
1168.63	1598.85
1296.92	1606.00
1633.82	

α-lactalbumin	
1091.36	1894.11
1200.65	2005.10
1204.61	2592.90
1669.00	3686.94
1700.12	4274.16
1780.29	4713.41

lysozyme	
606.0	1325.76
874.42	1333.77
893.39	1428.65
993.45	1675.46
1030.39	1753.84
1045.50	1803.70

myoglobin	
684.37	1518.00
735.28	1606.77
748.27	1661.56
790.00	1816.10
941.62	1853.90
1086.29	1857.00
1102.34	1885.00
1271.66	1936.79
1360.94	1982.06
1378.00	2109.97
1502.42	

β-lactoglobulin	
837.48	1193.49
853.64	1249.56
903.72	1635.63
933.68	1715.81
1121.35	

BSA	
665.40	1463.55
689.31	1464.68
712.38	1479.75
820.36	1502.57
847.54	1532.67
898.69	1554.56
927.74	1567.74
974.67	1576.78
975.60	1640.11
1068.50	1673.79
1072.43	1725.13
1107.45	1749.94

BSA (cont...)	
1108.15	1795.98
1138.43	1881.11
1163.65	1901.62
1166.53	1907.88
1193.65	1927.66
1195.84	2020.78
1249.73	2046.31
1271.67	2114.85
1284.15	2148.08
1306.03	2249.04
1399.65	2542.09
1419.63	2597.71
1439.90	2613.28
1443.70	

lactoferrin	
888.73	1362.60
946.41	1380.75
980.25	1381.70
990.00	1344.00
991.00	1233.00
1097.45	1475.83
1222.00	1490.67
1223.00	1505.97
1248.96	1551.00
1265.41	1592.05
1289.55	1674.84
1303.54	2241.57
1317.75	2279.5
1320.67	2315.47
1344.00	2373.24
1354.65	2559.39
1360.66	2687.99

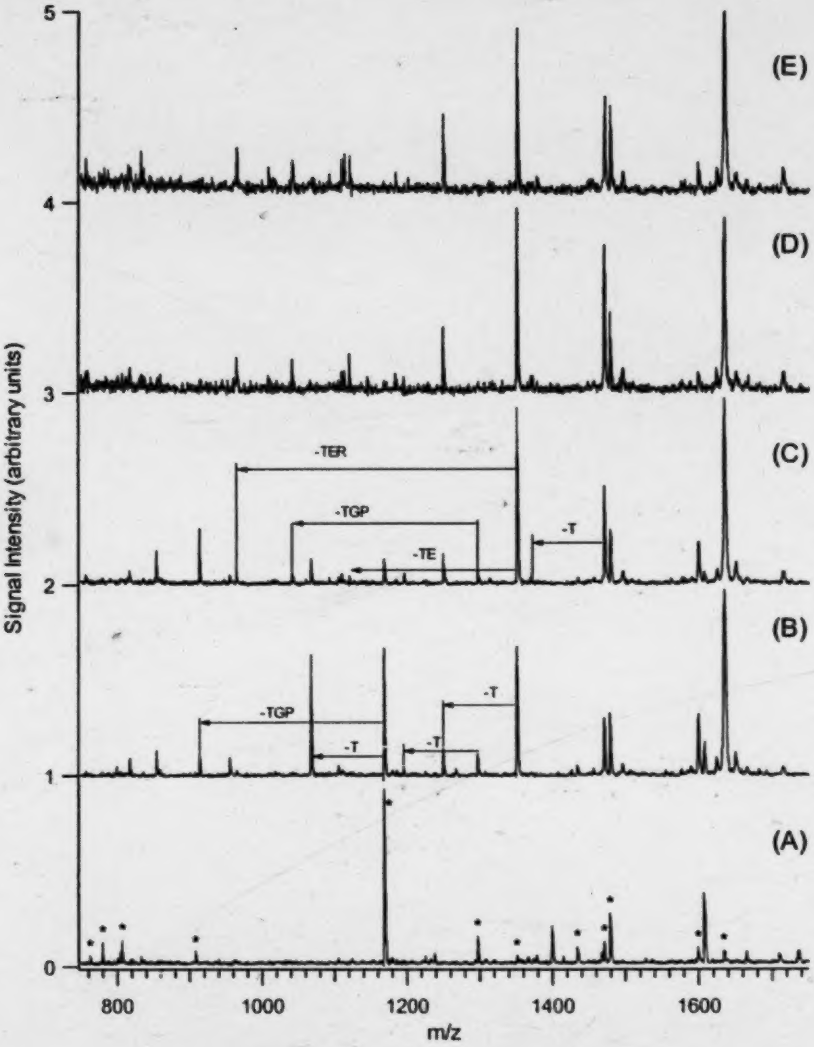

Figure 5.1 (A) MALDI mass spectrum of tryptic peptides from a digestion of 1 μg cytochrome c. The peaks labeled with an asterisk correspond to peptide fragments of cytochrome c. The tryptic peptides were then further digested with leucine aminopeptidase M (LAP) and MALDI mass spectra were recorded at various LAP digestion times: (B) 2 minutes, (C) 10 minutes, (D) 30 minutes, and (E) 60 minutes. The arrows indicate the new peaks that appear as a result of LAP digestion, their parent fragment, as well as the amino acid(s) cleaved.

The results of these digestions show that, under the conditions used, most of the more intense peaks appearing in the mass spectra can be assigned to peptide fragments of the digested protein, as we can see in Figure 5.1A. Between 7 and 52 peaks were identified as products of trypsin cleavage for each of the eight protein standards used.

In the sequential digestion procedure, LAP was applied directly to the tryptic peptides. N-terminal digestion of peptide fragments following trypsin digestion is preferred to C-terminal digestion (*e.g.*, using carboxipeptidases), since the C-terminus of trypsin-generated peptides would be either Lysine or Arginine. As a result, C-terminal amino acid sequencing would provide less additional information than N-terminal sequencing. The endoprotease, LAP, couples well with trypsin since the most efficient operating pH for LAP (7.0 to 9.0) matches that of trypsin (8.0). A solution of 0.1 M ammonium bicarbonate can therefore be used to buffer each digestion. Like trypsin, LAP operates most efficiently at 37°C, but is also active at room temperature. As well, LAP is capable of cleaving all N-terminal residues having a free α-amino or α-imino group, except for an X-proline sequence, where X represents a bulky hydrophobic residue [25]. Thus, we could theoretically expect this enzyme to create a ladder sequence for each peptide fragment generated by a trypsin digestion of a protein.

Table 5.2 summarises the results of N-terminal digestion of the peptide fragments following trypsin digestion. Several points are worth discussing from this table. First, it is noted that only a partial sequence tag of any particular peptide fragment can be obtained using LAP. The maximum observed number of amino acids cleaved from a single peptide is 4, in the case of the β-lactoglobulin parent peptide fragment (from trypsin digestion) at m/z 1715.8 Da. More typically, either 1 or 2 amino acids are cleaved from any given peptide.

When the parent masses listed in Table 5.2 are compared to the list given in Table 5.1, it is clear that not all peptides present in the solution were digested by LAP. Typically, amino acid tags are obtained for approximately one third of the tryptic fragments. Replicate digests of the samples indicate that the LAP will digest the same tryptic fragments, and with similar efficiencies. One possible explanation for this selective digestion is that LAP preferentially digests some amino acid residues over others. Previous studies have shown that the rate of hydrolysis depends on the polarity of the α-amino acid side chain [26]. These studies showed that the highest rate of hydrolysis of amides, anilides and β-naphthylamides was attributed to hydrophobic residues, in the order leucine>phenylalanine>valine>alanine>glycine [25]. However, to our knowledge, the relative rate of cleavage of amino acids in a peptide mixture has not been studied. For the protein samples used in this study, the aliphatic amino acid, leucine, was the most frequently digested, where 15 of the 43 leucine residues (35%) found at either the 1st N-terminal, or 2nd N-terminal position were digested. This compares to the average of 62 cleaved amino acids out of a total of 300 amino acids at the 1st or 2nd N-terminal position (21%). Other frequently digested amino acids include glycine (9 out of 20), glutamine (3 out of 8) and valine (5 out of 12). Thus, it appears that for a peptide mixture, LAP digests hydrophobic (non-polar) amino acids with higher efficiency. Several amino acid residues appearing at the 1st and 2nd N-terminal position were not cleaved by LAP. These include cysteine, aspartic acid, histidine, methionine, tyrosine, arginine and proline. The charged amino acid residues (Asp, Lys, Arg, Glu) appear to be cleaved with less efficiency (8.5% of first 2 residues) than the average rate of cleavage (21%).

Table 5.2 Peptide fragment products of LAP digestion following trypsin digestion of protein samples. Also indicated are the times, in minutes, when the peaks appear in the MALDI spectra, and the identity of the amino acid(s) cleaved.

parent mass (Da)	observed peak (time/min)	mass difference (amino acid(s) cleaved)
lysozyme		
993.46	806.99 (2)	186.47 (W)
1045.50	887.58 (10)	157.92 (GT)
	988.41 (2)	57.08 (G)
1325.76	1268.78 (2)	56.98 (G)
1428.65	1281.74 (2)	146.91 (F)
1675.46	1376.71 (2)	298.75 (IVS)
	1463.41 (2)	212.05 (IV)
	1562.31 (2)	113.15 (I)
1753.84	1639.81 (2)	114.03 (N)
ubiqiutin		
1067.69	851.63 (30)	216.06 (ES)
	938.87 (30)	128.82 (E)
1523.78	1167.59 (90)	356.19 (IQD)
	1282.35 (60)	241.43 (IQ)
	1410.66 (10)	113.12 (I)
myoglobin		
1271.66	1158.31 (2)	113.35 (L)
1606.77	1308.06 (10)	298.71 (VEA)
	1507.56 (2)	99.21 (V)
1661.56	1548.00 (10)	113.56 (L)
1816.10	1558.82 (10)	257.28 (GLS)
	1645.76 (10)	170.34 (LG)
1936.79	1676.25 (10)	260.54 (LF)
	1823.41 (2)	113.38 (L)

parent mass (Da)	observed peak (time/min)	mass difference (amino acid(s) cleaved)
lactoferrin		
888.73	801.21 (2)	87.52 (S)
946.41	859.35 (2)	87.06 (S)
1248.96	935.46 (2)	313.50 (LLS)
	1135.56 (2)	113.40 (L)
1317.75	1204.73 (2)	113.02 (L)
1320.67	1086.93 (2)	233.74 (SF)
	1233.54 (2)	87.13 (S)
1360.66	1161.73 (2)	198.93 (KA)
1362.60	1218.65 (2)	143.95 (GS)
	1305.51 (2)	57.09 (G)
1381.70	1154.71 (2)	226.99 (QV)
	1253.72 (2)	127.98 (Q)
1592.05	1462.45 (2)	129.60 (E)
cytochrome c		
1168.63	913.67 (2)	254.96 (TGP)
	1067.46 (2)	101.11 (T)
1296.92	1041.43 (10)	255.49 (TGP)
	1195.63 (2)	101.29 (T)
1350.99	964.49 (10)	386.50 (TER)
	1120.64 (10)	230.35 (TE)
	1249.79 (2)	100.94 (T)
1470.69	1370.04 (2)	100.65 (T)

parent mass (Da)	observed peak (time/min)	mass difference (amino acid(s) cleaved)
BSA		
1163.63	1050.74 (2)	112.89 (L)
1249.62	1102.64 (2)	146.98 (F)
1479.80	1367.15 (2)	112.65 (L)
	1308.17 (2)	171.63 (LG)
1639.94	1511.90 (2)	128.04 (K)
1674.75	1546.42 (10)	128.33 (Q)
β-lactoglobulin		
837.48	766.66 (2)	70.82 (A)
853.64	782.38 (2)	71.26 (A)
1193.49	981.65 (45)	211.84 (VL)
1715.81	1254.45 (45)	461.36 (LSFN)
	1368.70 (2)	347.11 (LSF)
	1515.49 (2)	200.32 (LS)
	1602.40 (2)	113.41 (L)
α-lactalbumin		
1091.36	978.32 (17)	113.04 (L)
1200.65	931.35 (17)	269.30 (VGI)
	1044.52 (2)	156.13 (VG)
	1101.48 (2)	99.17 (V)
1700.12	1439.42 (17)	260.70 (FL)
	1552.57 (17)	147.55 (F)
1780.29	1595.60 (2)	184.69 (AL)
	1708.81 (2)	71.48 (A)

It is obvious that experimental parameters such as molar enzyme to substrate ratio and digestion time can affect the outcome of the sequential digestion method. Using cytochrome c as an example, the ratio of LAP to substrate can be estimated by assuming that trypsin quantitatively cleaves the protein into tryptic fragments with molecular weights ranging from 500 to 2000 Da. We would therefore have a total of about 1 nmol of tryptic peptides. Adding 0.02 nmol LAP (MW 280 kDa) to this solution resulted in an approximate 1:50 ratio of enzyme to substrate. We would therefore expect to see the cleavage products resulting from the most favourable LAP substrate appearing at earlier times in the spectrum. The progression of the digestion through time is depicted in Figure 5.1A-E. This figure shows the MALDI mass spectra obtained from 1 μg cytochrome c that was first digested with trypsin and then with LAP for varying amounts of time. Figure 5.1B shows that 2 minutes after the LAP was added to the peptide sample, several new peaks resulting from single amino acid cleavage are evident. If we allow this reaction to proceed for 10 minutes, we can observe even more peaks resulting from N-terminal digestion (Figure 5.1C). As the reaction progresses to longer times, no new cleavage products are evident in the MALDI spectra. However, when other protein samples were subjected to the sequential digestion procedure, new peaks could be observed well beyond 10 minutes digestion by LAP. As an example, Figure 5.2 shows the sequential digestion of β-lactoglobulin at various digestion times up to 45 min. From this figure, we see that the intensity of the peak at m/z 1368, which corresponds to N-terminal cleavage of the first 3 amino acids (Leu, Ser, and Phe) from the peptide fragment of mass 1715.8 Da, is still increasing in intensity at 45 minutes LAP digestion, an indication that this product is still being formed at this time period. More conclusive,

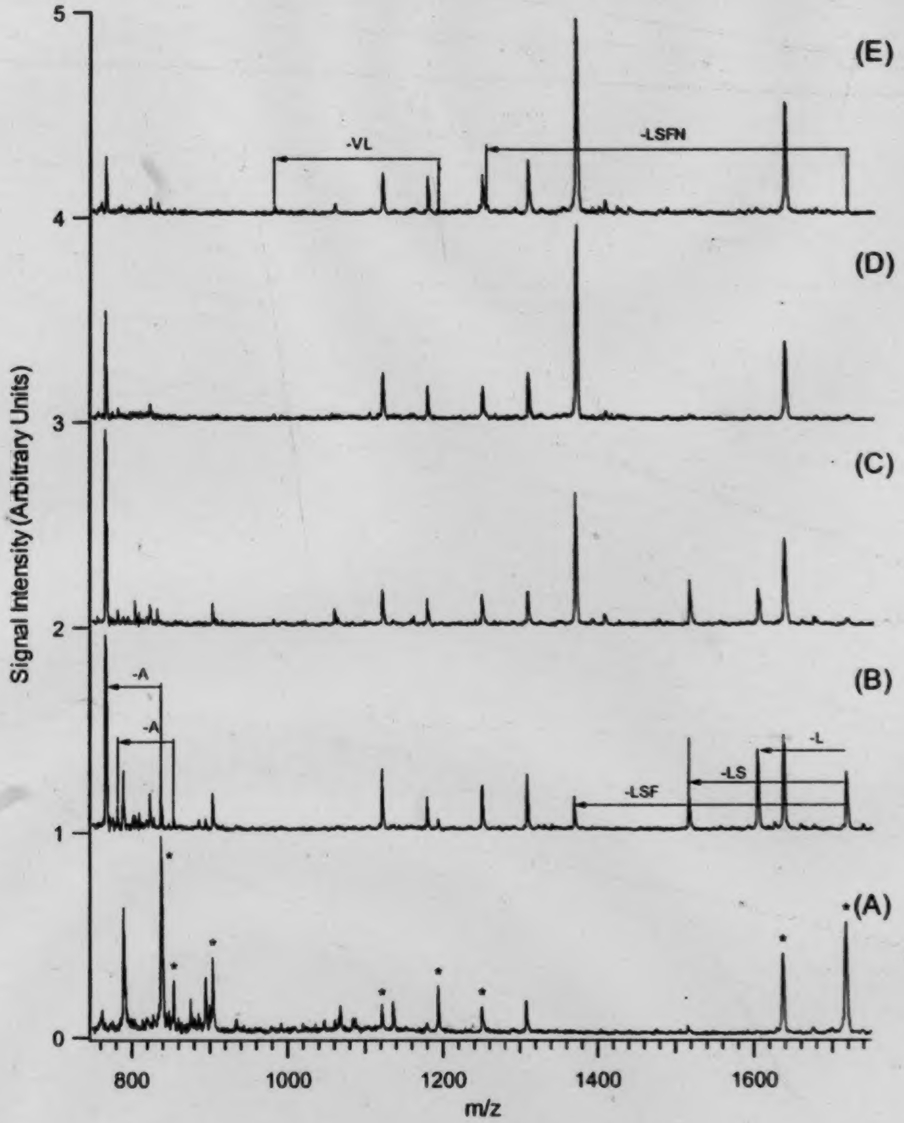

Figure 5.2 (A) MALDI mass spectrum of tryptic peptides from a digestion of 1 μg of β-lactoglobulin. The peaks labeled with an asterisk correspond to peptide fragments of β-lactoglobulin. The tryptic peptides were then further digested with leucine aminopeptidase M (LAP) and MALDI mass spectra were recorded at various LAP digestion times: (B) 2 minutes, (C) 10 minutes, (D) 17 minutes, and (E) 45 minutes.

the peak corresponding to the cleavage of the fourth amino acid residue, Asn, from the same peptide fragment is only observed in the MALDI spectrum after 45 minutes digestion with LAP. As another example (spectra not shown), when the tryptic peptides of ubiquitin were digested with LAP, the peptide fragment at 1524 Da loses the first amino acid, Gln, resulting in a peak at 1411 Da in the spectrum at 10 minutes. The fragment resulting from N-terminal cleavage of the second amino acid, Ile, is only observed in the spectrum at 60 minutes digestion, and the fragment generated by the cleavage of the third amino acid, Asp, is first seen 90 minutes after LAP was first added to the protein sample.

Although most of the cleavage products result in the first 10 minutes of digestion with LAP (see Table 5.2), it is clear that some additional cleavage products are generated as the reaction proceeds beyond 10 minutes. Since only a small portion of the total digestion needs be sampled for MALDI analysis, a given digestion can be analysed at several time intervals. Also, in analyzing a digestion at various time intervals, the ability to overlay several spectra, as in Figure 5.1, simplifies the task of picking out new peaks resulting from LAP digestion. Those peaks with signal to noise ratios that are only slightly above the background are still confidently observed, as the detection of these peaks in multiple spectra serves to confirm their presence.

The observation that new cleavage products are formed at much later times following the addition of LAP to the sample leads one to assume that if the digestion time is extended further, the digestion will continue to proceed. However, in practice, this was not observed. As the reaction time was extended, an overall degradation in the MALDI signal was observed. A decrease in signal intensity was observed immediately upon addition of LAP to the sample, likely a result of the high salt concentrations of $MgCl_2$ and

$(NH_4)_2SO_4$ in the LAP slurry. However, the signal intensity continued to decrease as the digestion proceeded to later times. We can easily see the signal degradation in Figure 5.1 as the reaction proceeded. One possible explanation for this is that as the N-terminal digestion proceeds, the cleaved amino acid residues resulting from this digestion would increase in concentration, creating a suppression effect on the ionization of the remaining peptide fragments. Also, the total intensity of a given peptide fragment in the mass spectrum would be split between the intact peptide fragment and those resulting from cleavage of the first, second, third, or subsequent amino acid from the peptide fragment. These cleaved fragments may not be present at concentrations high enough to be observed in the mass spectrum. Although the digests were analysed by MALDI for up to 24 hours following LAP addition to the sample, new digestion products were never observed beyond 90 minutes digestion. Therefore the length of the peptide ladder sequences are generally limited to short tags of 2 to 3 amino acids.

The above results indicate that the sequential digestion procedure will not generate ladder sequences for all tryptic peptides, and therefore cannot always be used to sequence a specific peptide present in a digested sample. However, the ability to generate additional peptide fragments for about 1/3 of the tryptic digest does provide additional information that can aid in protein identification (see below). It is also clear that with a fixed enzyme to substrate ratio sampling the resulting digests from LAP digestion of tryptic peptides at different time intervals are advantageous in revealing the amino acid sequence information.

5.3.2 In-Gel Digestion. Proteome separation is usually accomplished with gel electrophoresis. The performance of the sequential digestion procedure for in-gel digestion was investigated. For in-gel digestion, 4 of the 8 proteins were selected for this study. 1 μg of each protein sample was loaded onto the gel and subjected to electrophoresis. The bands were excised from the gel, ground to small pieces; disulfide bonds were reduced and the cysteines were blocked to prevent the disulfide bonds from reforming. The protein samples were then digested with trypsin and the fragments were extracted. The results are summarized in Table 5.3. We can see that most of the peaks identified after trypsin digestion of 1 μg protein (see Table 5.1) could also be observed in the MALDI spectra of the gel-separated proteins. The MALDI spectra of the proteins from the gel were not as intense as those out of gel, which can be explained by lower in-gel digestion efficiency, lower extraction efficiency and an increase in contaminants resulting from the gel digestion protocol.

N-terminal digestion of these samples was performed after the samples had been extracted from the gel. By adding the LAP to the already extracted peptides, the contact between peptide and enzyme could be maximized, as the extracted peptides would be free from the gel matrix and concentrated in solution following evaporation of the extracting solvent. Also, LAP digestion of the extracted tryptic peptides can facilitate the time-course experiment. The reaction time can be precisely controlled as the reaction can be immediately stopped by adding the sample to the matrix solution. As before, 1 μl of LAP (5.6 μg) was used to digest the peptide fragments in each sample. These results are also shown in Table 5.3. Again, results were similar to those obtained from the solution-phased digests. Some peaks that were observed in the solution-phased digests were no longer observed for the gel-separated proteins. This may be a result of the weaker overall intensity of the MALDI peptide spectra for the gel-separated proteins or the lower

Table 5.3 Peptide fragment products of LAP digestion following trypsin digestion of 1 μg protein samples that were extracted from polyacrylamide gels. Also indicated are the times, in minutes, when the peaks appear in the MALDI spectra, and the identity of the amino acid(s) cleaved.

parent mass (Da)	new peak mass (time/ min)	mass diff. (amino acid(s) cleaved)
lysozyme		
606.26	-	-
674.42	-	-
993.56	-	-
1030.67	-	-
1045.50	887.46 (2) 988.54 (2)	158.04 (GT) 56.96 (G)
1325.70	-	-
1333.71	-	-
1428.65	1281.79 (15)	146.86 (F)
1491.25	-	-
1675.33	1377.73 (2) 1463.50 (15)	297.60 (IVS) 211.83 (IV)
1753.84	1639.57 (15) 1538.37 (65)	114.27 (N) 215.47 (NT)
myoglobin		
684.37	-	-
735.27	-	-
748.26	-	-
1271.66	1158.49 (2)	113.17 (L)
1360.82	-	
1479.53	1178.1 (2) 1250.14 (2) 1378.24 (2)	301.43 (TEA) 229.39 (TE) 101.29 (T)

parent mass (Da)	new peak mass (time/ min)	mass diff. (amino acid(s) cleaved)
myoglobin (cont...)		
1502.52	-	-
1606.60	1307.95 (2) 1507.85 (2)	298.65 (VEA) 98.75 (V)
1815.67	-	-
1853.85	-	-
1937.02	1677.34	259.68 (LF)
1981.86	-	-
β-lactoglobulin		
837.48	766.84 (2)	70.64 (A)
853.61	782.80 (2)	70.81 (A)
903.68	-	-
916.61	-	-
1065.78	-	-
1193.68	-	-
1635.5	-	-
1715.81	1253.94 (30) 1368.06 (2) 1515.14 (2) 1602.33 (2)	461.87 (LSFN) 347.75 (LSF) 200.67 (LS) 113.48 (L)

parent mass (Da)	new peak mass (time/ min)	mass diff. (amino acid(s) cleaved)
BSA		
689.77	-	-
847.89	-	-
898.92	-	-
928.02	-	-
1138.83	-	-
1163.97	1050.46 (2)	113.51 (L)
1193.91	-	-
1250.15	974.14 (2) 1102.08 (2)	276.01 (FK) 148.07 (F)
1306.16	-	-
1308.29	-	-
1420.1	-	-
1480.07	1366.92 (2)	113.15 (L)
1490.92	-	-
1567.74	-	-
1639.94	1511.46 (2) 1412.67 (30)	128.48 (K) 227.27 (KV)
1749.11	-	-
1881.51	-	-
2021.58	-	-

concentration of peptides present following gel extraction. On the other hand, some N-terminal digestion products were visible in the MALDI spectra after gel separation that had not been detected in the free solution experiment. This observation may be attributed to the variation in the ratio of LAP to substrate, as the peptide concentration is altered due to difference in digestion and extraction efficiencies in the gel.

5.3.3 Sensitivity. In order to assess the sensitivity of this method for cleaving very small amounts of starting protein sample, 100 ng of lysozyme (7 pmol) was digested with 50 ng trypsin in 7 μl total sample volume. The lysozyme digest was then diluted to a total volume of 14 μl with 0.1 M NH_4HCO_3, thus forming a 0.5 μM lysozyme solution. A 1-μl portion of the diluted lysozyme digest was added to 4.5 μl matrix solution and analyzed by MALDI-TOF-MS. A 6 μl portion of the diluted lysozyme digest (3 pmol) was then digested with 0.5 μl of diluted LAP (1.4 μg). After allowing the LAP to digest for various times, 1 μl was mixed with 4.5 μL matrix solution and 0.5 μL was spotted on the MALDI target for analysis, representing a total sample loading of 42 femtomoles. From this MALDI spectrum, 10 of the 12 peaks that matched the theoretical lysozyme digest using 1 μg lysozyme could still be observed here. When the sample was digested with LAP, 7 of the 9 N-terminal digestion products that were observed from the LAP digest of 1 μg lysozyme could be seen here. All of the LAP digestion products could be seen after 2 minutes digestion with LAP. No new peaks were observed as the digest was allowed to proceed. This is consistent with what was observed for the 1 μg digestion.

100 ng BSA was also digested with 50 ng trypsin in 7 μl total digestion solution. The digest was diluted to 14 μl with 0.1 M NH_4HCO_3, representing approximately 100 nM BSA. 1 μl of the sample was added to 4.5 μl matrix solution analyzed by MALDI

TOF-MS. 3.5 μl was then digested with 0.5 μl of twice diluted LAP (1.4 μg). Each MALDI spectrum was collected from a total sample loading of 9 femtomoles on the MALDI probe tip. From the trypsin digest spectrum, 11 of the 20 peaks originally identified using 1 μg BSA could still be observed. LAP digestion revealed 3 of the 6 new peaks resulting from N-terminal cleavage, all within the first 2 minutes digestion.

For samples loaded onto SDS-polyacrylamide gels, it was found that 100 ng of sample loaded onto the gel could still produce visible N-terminal cleavage products when subjected to the digestion protocol. The results of this study are summarized in Table 5.4. The MALDI mass spectra collected for the digestion of lactoferrin are shown in Figure 5.3 (A through C). It is evident from the trypsin digest (Figure 5.3A) that the signals observed are weaker, and that less peptide fragments are detectable. However, even at these low levels, some N-terminal digestion products are observable.

From these experiments, we can conclude that N-terminal digestion of trypsin-digested proteins is sensitive at the sub-picomole level, which is similar to what was reported for an on-probe digestion procedure [19]. However, as lower starting amounts of sample are used, the amount of observable cleavage products in the MALDI spectrum is further reduced. Although the volume of starting material can still be lowered, it is the concentration of protein in solution that ultimately limits the success of this procedure.

Table 5.4 Peptide fragment products of LAP digestion following trypsin digestion of 100 ng protein samples that were extracted from polyacrylamide gels. Also indicated are the times, in minutes, when the peaks appear in the spectra, and the identity of the amino acid(s) cleaved.

parent mass (Da)	new peak mass (time/ min)	mass diff. (amino acid(s) cleaved)	parent mass (Da)	new peak mass (time/ min)	mass diff. (amino acid(s) cleaved)	parent mass (Da)	new peak mass (time/ min)	mass diff. (amino acid(s) cleaved)
cytochrome-c			**BSA**			**lactoferrin**		
805.67	-	-	906.52	-	-	806.94	-	-
907.61	-	-	917.36	-	-	831.75	-	-
1168.57	1067.38 (2)	101.19 (T)	927.49	-	-	1063.86	-	-
1297.24	1195.88	101.36 (T)	1082.31	-	-	1097.81	-	-
1350.59	1249.81	100.78 (T)	1153.45	-	-	1135.81	-	-
1433.35	-	-	1168.46	-	-	1202.15	-	-
1478.69	-	-	1197.38	-	-	1265.56	-	-
			1249.25	-	-	1303.74	-	-
			1308.02	-	-	1317.98	1204.18 (2)	113.80 (L)
			1362.99	-	-	1320.89	1234.15 (2) 1086.13 (2)	86.74 (S) 234.76 (-SF)
			1479.91	1366.81 (10)	113.19 (L)	1355.07	-	-
			1490.91	-	-	1460.22	-	-
			1568.43	-	-	1473.81	-	-
			1640.1	-	-	1490.74	1362.59 (2) 1305.44 (2)	128.15 (K) 185.30 (KG)
						1593.04	-	-
						1675.32	-	-
						2372.03	-	-

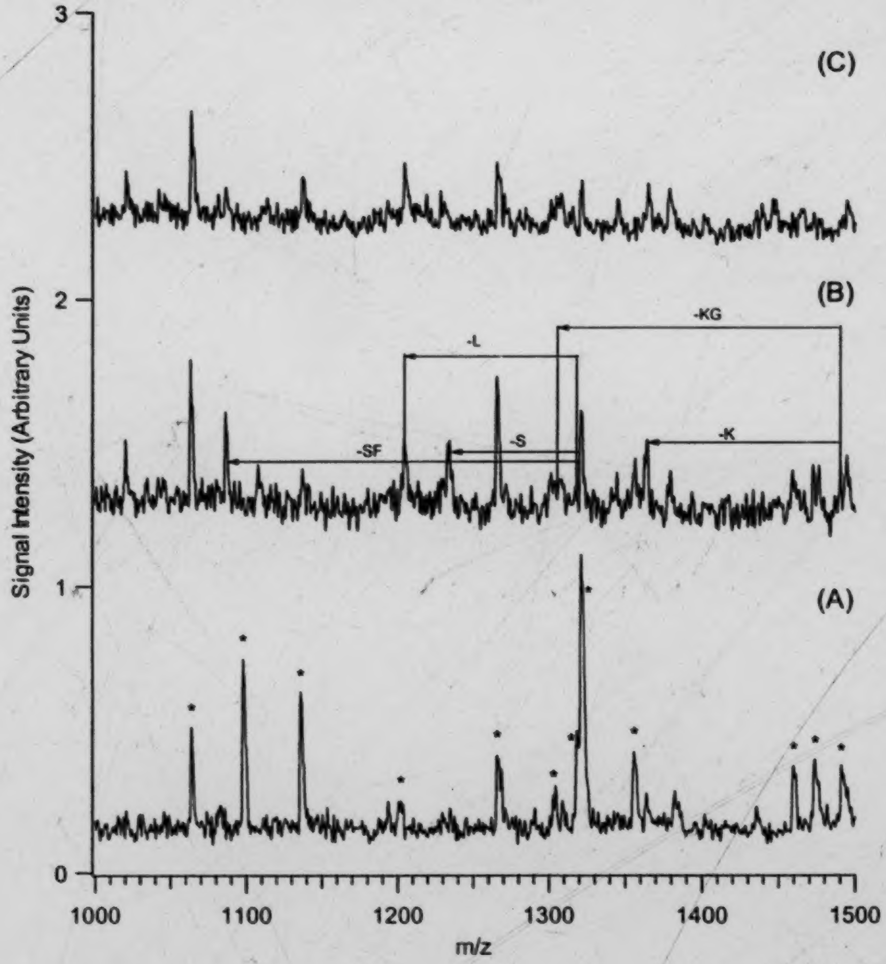

Figure 5.3 (A) MALDI mass spectrum obtained from trypsin in-gel digestion of 100 ng of lactoferrin with the lactoferrin peptides labeled with an asterisk. (B) MALDI mass spectrum obtained after 2 minutes LAP digestion of the tryptic peptides. (C) MALDI mass spectrum obtained after 10 minutes LAP digestion of the tryptic peptides.

5.3.4 *E. coli* Fractions. The use of the sequential digestion procedure to increase the confidence of protein identification based on mass fingerprinting is illustrated in the example of identifying bacterial proteins. An *E. coli* extract was fractionated by HPLC and the fractions were subjected to MALDI analysis. Two fractions were chosen to demonstrate the applicability of the digestion protocol.

The first fraction contains a strong peak at m/z 7870.0 Da, along with two other peaks shifted 16 and 32 Da up from 7870 (see Figure 5.4). The doubly and triply charged peaks are also observed in the spectrum. It is evident from the spectrum that the HPLC separation did not result in a pure, single component fraction. Several peaks are seen in the low mass region of the spectrum between 2000 and 5000 Da. These peaks are most likely from low molecular weight proteins. The relative amounts among the proteins in the fraction are unknown, since MALDI signals in the spectrum are not quantitative.

The fraction was subjected to the digestion protocol described in the experimental section. Figure 5.5 displays the results of these digestions. From Figure 5.5A, we can see that a very clean trypsin digestion results, despite the large number of high intensity low mass peaks seen in the MALDI spectrum of the undigested fraction. The masses of 29 peaks in Figure 5.5A were entered into the PeptIdent program, searching the SWISS-PROT database, which contained about 80,000 total entries, under the subset of species *Escherichia coli*, between the mass range 0 to 10000 Da. Several possible proteins were found, with the highest score owing to the 50S ribosomal protein L31 (AC # P02432). Eight of the entered peptides matched this protein, giving sequence coverage of 85.7%. The protein has a molecular weight of 7871.06; however, up to 2 disulfide bonds could be present due to the presence of four cysteine residues in the protein, which would shift this mass down by either 2 or 4 Da.

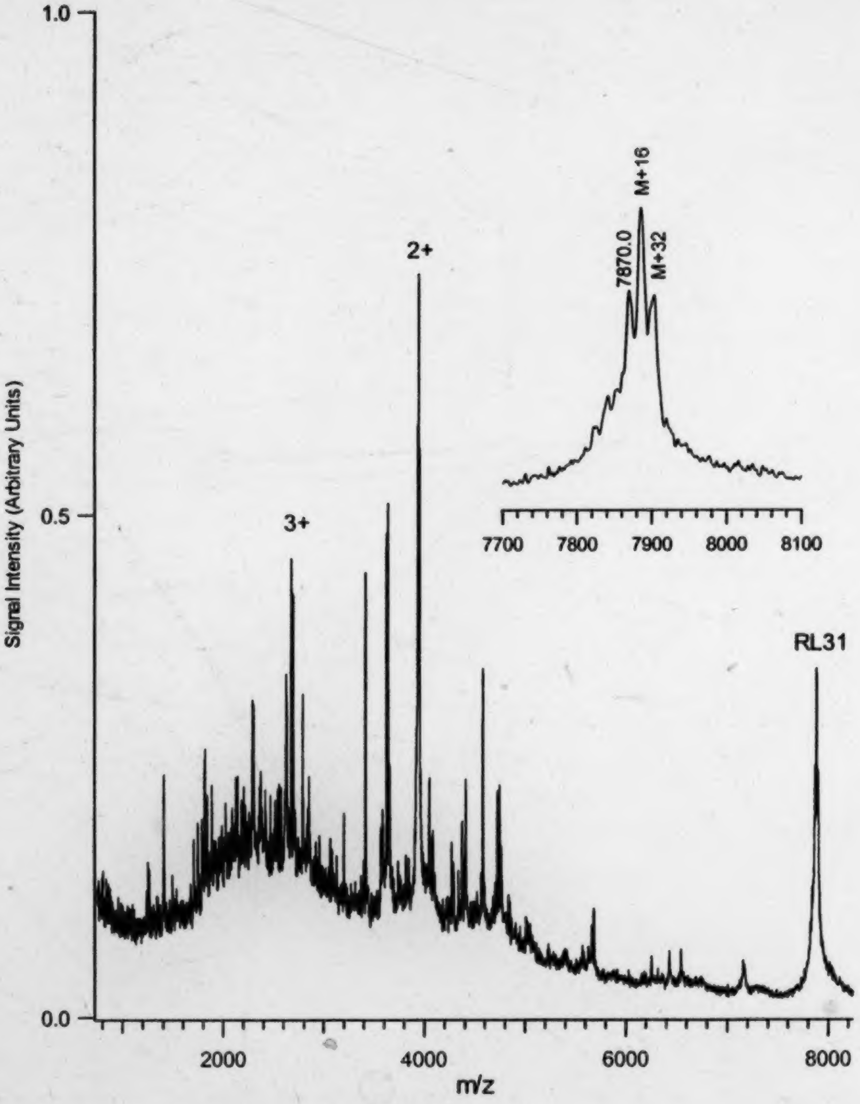

Figure 5.4 MALDI mass spectrum of a HPLC fraction from *E. coli* cell extract, showing an intense peak at mass 7870 Da. An expanded view of this peak is shown in the insert.

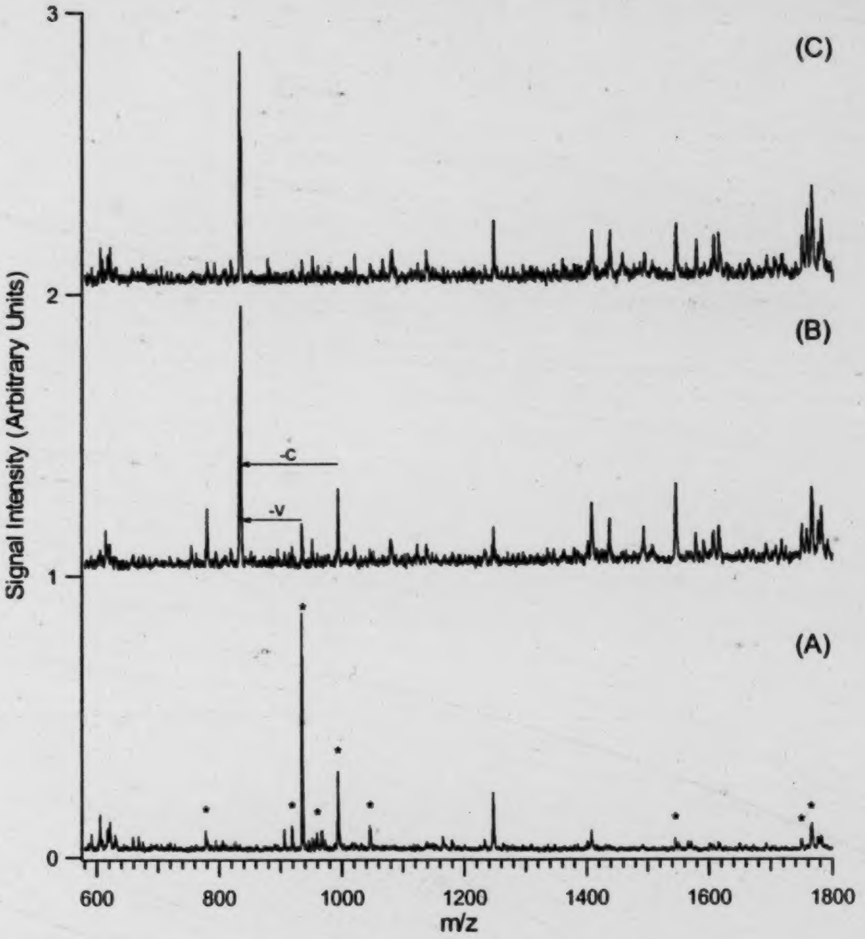

Figure 5.5 (A) MALDI mass spectrum obtained from trypsin digestion of the HPLC fraction shown in Figure 5.4. (B) MALDI mass spectrum recorded after 2 minutes LAP digestion of the tryptic peptides. (C) MALDI mass spectrum recorded after 10 minutes LAP digestion.

The protein also contains two methionines, and thus the peaks shifted up 16 and 32 Da from 7870 can be attributed to methionine oxidation. As shown in Figure 5.5B, the N-terminal digestion gave two strong peaks, at 833.1 Da and 835.2 Da, which are attributed to the loss of the first amino acid from the peaks at 934.4 (-V) and 993.5 (-C) respectively. The combination of protein molecular weight, trypsin digestion, with 85.7% sequence coverage, and the two N-terminal digestion products unambiguously identifies this protein as *E. coli* 50S ribosomal protein L31.

A second fraction was chosen to illustrate the identification of two proteins in the same sample by the sequential digestion procedure. The MALDI spectrum of this fraction, shown in Figure 5.6, indicates the presence of two protein components in the fraction. The first protein has molecular weight 5097.6 Da, and the second is located at m/z 5382.1. Peaks shifted up +16 and +32 Da are also seen in the spectrum. The doubly charged species are also shown. As with the previous fraction, several other peaks are also seen in the spectrum, which could pose interference problems following trypsin digestion. The trypsin digestion, along with the LAP digestion at various times is shown in Figure 5.7. A total of 54 peaks were detected in the trypsin digest spectrum between mass 500 and 2500 Da. These peaks were entered into the PeptIdent program, searching the SWISS-PROT Database with 79909 entries, looking at the subset species *Escherichia coli* in the mass range from 0 to 11000 with a minimum 3 peptides to match. The results of this search gave 119 possible matches, with the highest score owing to 50S ribosomal protein L27, having 10 peptides matching, and sequence coverage of 71.4%. Although N-terminal digestion of the sample produced several new peaks, no new peaks could be attributed to this 50S ribosomal protein. The protein that gave the next highest score, with 7 peptides, was examined. It is a hypothetical 30S ribosomal protein S22, having a mass of 5095.8 Da. When the N-terminal digestion products were fitted to a theoretical digestion of this protein, 3 peaks could be attributed to LAP digestion of RS22 trypsin peaks. This confirms the identity of this protein in the sample.

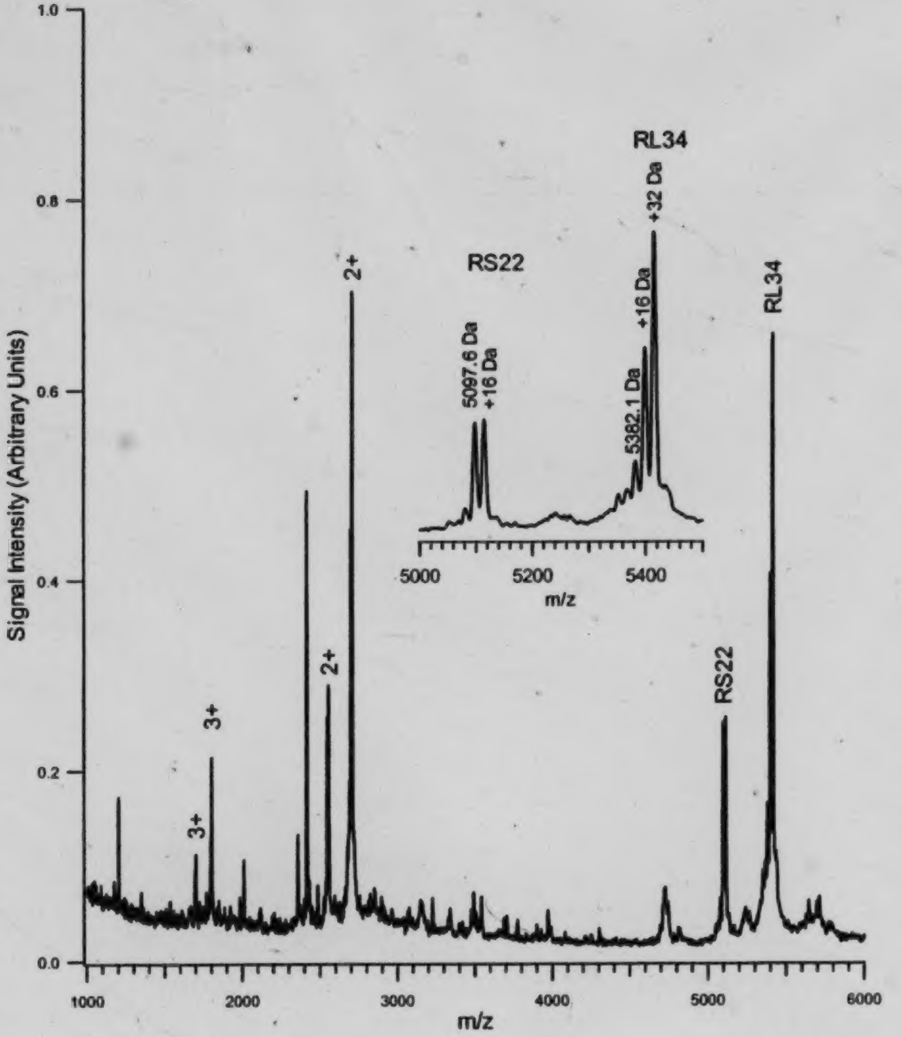

Figure 5.6 MALDI spectrum of a HPLC fraction from *E. coli* cell extract, showing two main protein components. An expanded view of the singly charged molecular ion peaks is shown in the insert.

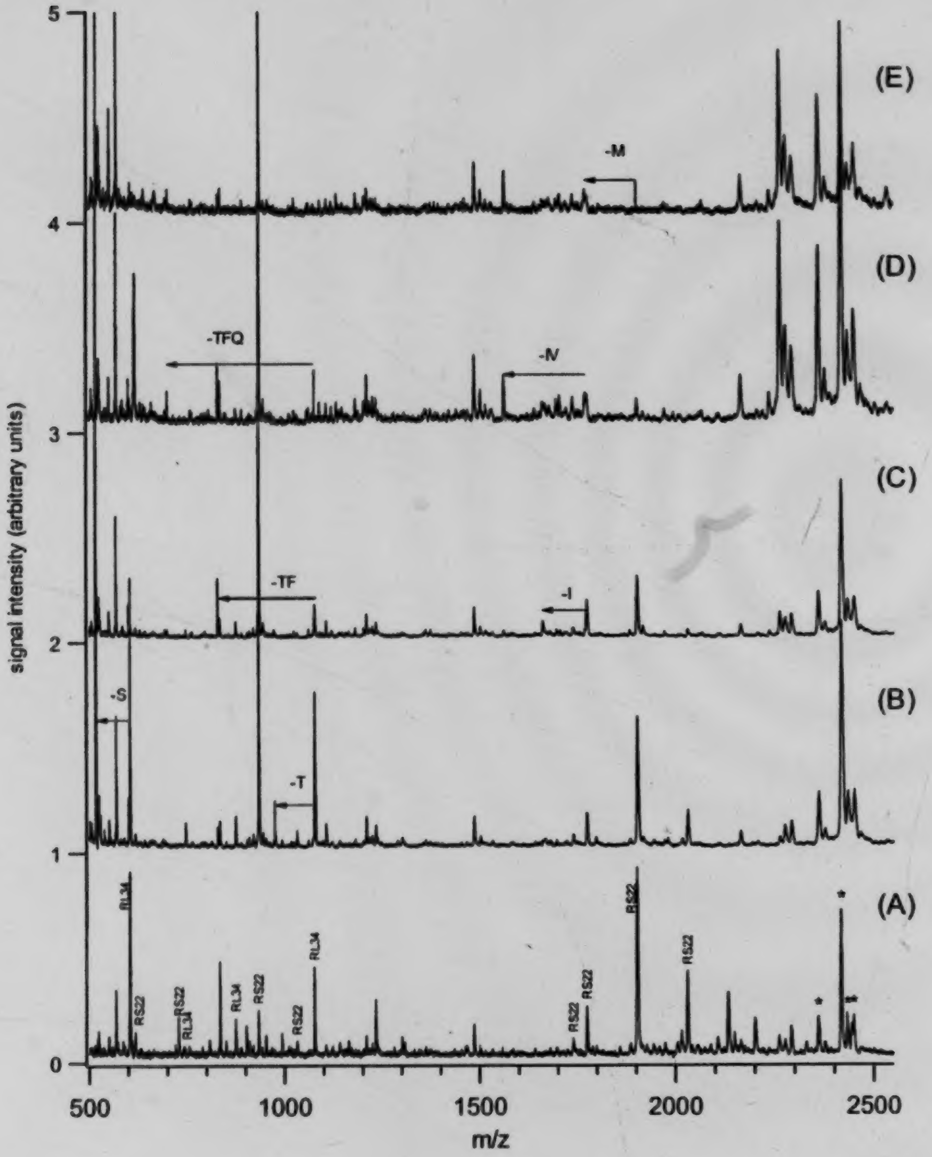

Figure 5.7 (A) MALDI mass spectrum obtained from trypsin digestion of the HPLC fraction shown in Figure 5.6. (B) MALDI mass spectrum recorded after 2 minutes LAP digestion of the tryptic peptides. (C) MALDI mass spectrum recorded after 10 minutes LAP digestion. (D) MALDI mass spectrum recorded after 30 minutes LAP digestion. (E) MALDI mass spectrum recorded after 60 minutes LAP digestion.

The second sample component was identified as the 30th highest scoring protein on the list of possible protein generated from the PeptIdent search, namely 50S ribosomal protein L34. This protein has a molecular weight of 5380.39, a difference of 318 ppm from the peak found in the MALDI spectrum of the fraction. Four peptides matched this protein, with sequence coverage of only 50% for this relatively small protein. However, the N-terminal digestion gave 4 new peaks, including a sequence tag of 3 amino acids from the peptide fragment at 1076.7 Da. In this case, N-terminal digestion was necessary to unambiguously identify this protein, as the trypsin digestion alone gave low sequence coverage, and the PeptIdent search gave so many other possible choices for the protein.

Other possible proteins from the list generated by the PeptIdent search may be present in the mixture, but their presence cannot be confirmed. For example, the 7th highest scoring protein on the list, with 6 peptides matching and sequence coverage of 53.2%, namely a hypothetical 10.5 kDa protein in the hyba-exbd intergenic region of E coli, may be present in the mixture. This protein has a mass of 10495.36 Da. A peak at 10490 Da is present in the MALDI spectrum (results not shown in the figure); however, LAP digestion did not produce any digestion products that could be matched to this protein and therefore, this protein could not be unambiguously identified.

5.3.5 CD9 Sample. The Raji/CD9 B-cell lymphocyte lysate was passed through an affinity column to capture CD9. The antigen was then eluted using 0.05 M diethylamine, pH 11.5, and acidified with 2% acetic acid. The fraction was separated by 1-dimensional SDS-PAGE, followed by Western blotting with mAb 50H.19.HRP and ECL/exposure on film. Western blotting confirmed the presence of CD9 in the fraction [21]. A second gel was stained with coomassie blue, and the suspected CD9 spot (based on Western blot from previous gel) was excised and digested with trypsin. Following digestion, the supernatant digestion buffer was mixed in a 1:1 ratio with matrix solution and spotted on a MALDI target for analysis. Based on the mass of the detected peaks, 7 peptide fragments of CD9 were identified. These seven fragments correspond to the amino acid sequence of position 119-143 and 169-178, which represents 19% coverage. These positions correspond to sections of the extracellular domain of the CD9 molecule. This hydrophilic region is most likely to be soluble in an aqueous fraction of the gel extract, as well as most likely to be ionized in the MALDI process.

The gel piece was further extracted with an acetonitrile/ aqueous TFA solution in order to solubilize more peptide fragments for MALDI analysis. The extract was concentrated by evaporating the solvent and reconstituting the sample in 3.3 μL of matrix solution. The MALDI spectrum revealed 6 peaks that corresponded to theoretical CD9 peptide fragments. Of these peaks, 5 were also observed in the aqueous extract. In addition, a peak of high intensity, corresponding to the sequence of position 222-227, was also observed in this extract. This fragment also corresponds to a hydrophilic domain of the CD9 molecule, representing a segment of the cytoplasmic domain of the protein. No peptides corresponding to the transmembrane segments of the CD9 molecule were observed in the MALDI spectra.

It is noted that a theoretical digest of CD9 would produce few peptide fragments, as the CD9 molecule contains very few lysine or arginine residues. Higher molecular weight fragments are very difficult to extract from the gel. In addition, the CD9 molecule has four hydrophobic segments (transmembrane) which further prevents solubilization of the larger peptide fragments from the gel. This therefore leads to a very incomplete peptide mass map of the CD9 molecule when digested with trypsin alone, giving a low confidence in identifying the sample.

To increase the confidence in identifying the gel spot as being CD9, further amino acid sequencing was attempted. In addition to the MALDI analysis of the trypsin digest, the supernatant digestion buffer was also analyzed by nanospray MS on a Bruker ion trap mass spectrometer. However, owing to the low concentration of peptides in the sample, this experiment failed to reveal any CD9 fragments. Also, post source decay (PSD) was attempted on the peptides observed in MALDI; however no PSD fragments were observed. The sample was therefore subjected to N-terminal sequencing with LAP.

A 0.4 μl portion of 0.2 μg/μl LAP was added to 10 μl of the supernatant digestion buffer solution (pH 8), which was then allowed to digest at room temperature for various times before a portion of the solution was removed and mixed with matrix to stop the digest. MALDI spectra were recorded following digestion with LAP for 2, 10, 30, and 60 minutes. The spectra obtained at 10 minutes and 30 minutes are displayed in Figure 5.8, along with the spectra corresponding to the trypsin digestion of the sample. One can observe from Figure 5.8 that the peptide fragment with mass 1344.7 Da (amino acid position 133-143) was digested by LAP, revealing new peaks at mass 1243.7 and 1115.5 Da. These peaks correspond to the loss of threonine (T) and lysine (K) from the N-terminus of the peptide fragment. These additional peaks serve to confirm the identity of the peak at m/z 1344.7, and therefore the identity of the CD9 protein.

This example shows that, for a low abundance hydrophobic protein where nanospray ESI MS/MS and post-source decay failed to produce any results, the sequential procedure can generate useful data to confirm the identity of the protein based on MALDI mass peptide fingerprinting.

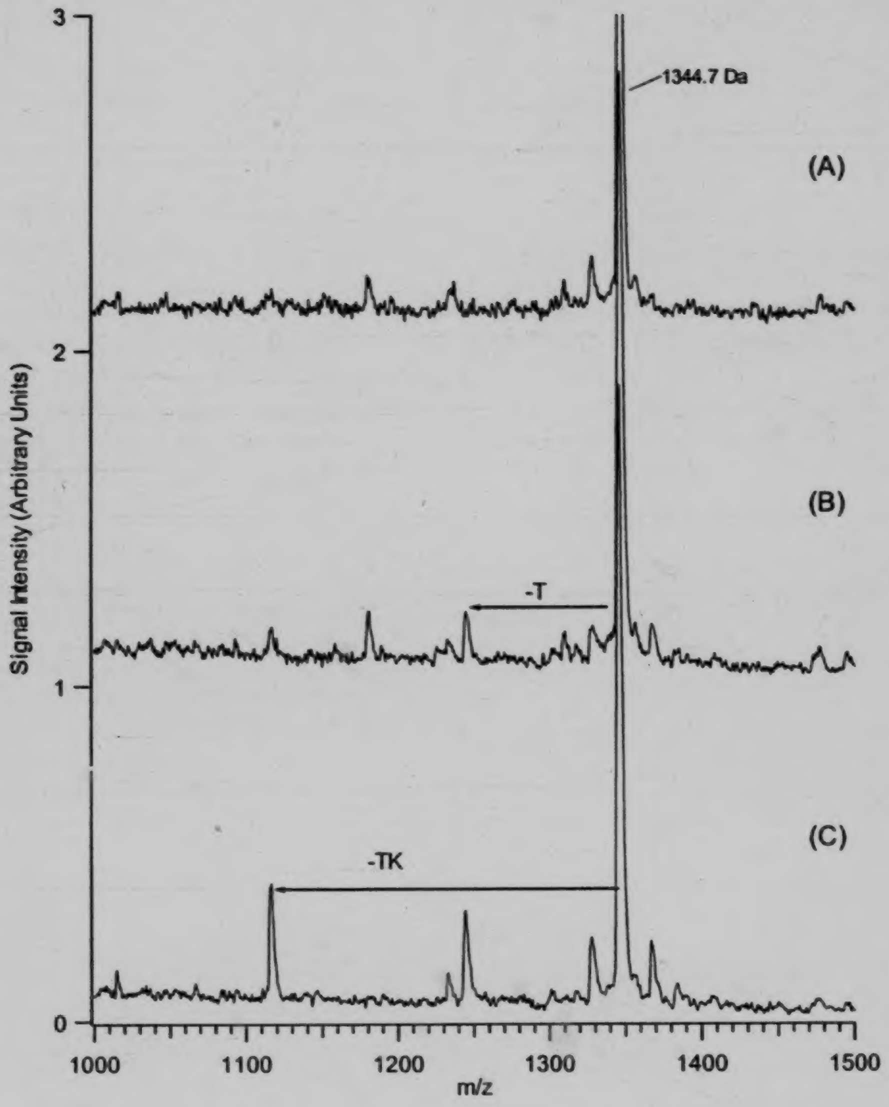

Figure 5.8 (A) MALDI mass spectrum obtained from trypsin in-gel digestion of CD9. (B) MALDI mass spectrum recorded after 10 minutes LAP digestion of the tryptic peptides. (C) MALDI mass spectrum recorded after 30 minutes LAP digestion.

The above examples illustrate that the sequential digestion procedure can provide additional information for unambiguous identification of proteins based on MALDI peptide mass mapping alone. It is a straightforward procedure and can be applied to identifying proteins in a simple mixture. We have used this procedure to identify a number of HPLC fractionated bacterial proteins. There are, of course, alternative methods for identifying these proteins or confirming the identity of the protein identified by peptide mass fingerprinting. Nanospray or nano-LC/ESI MS/MS could be used for identification and would involve less human intervention for data processing, which can be an important factor in determining the method to be used for proteome analysis.

The recent acquisition of a Q-TOF-MS instrument equipped with a MALDI source for CID fragmentation of peptides opens new opportunities for peptide amino acid sequencing through tandem MS fragmentation. However, peptide fragmentation in a Q-TOF may be difficult or impossible for high molecular weight peptides, owing to their difficulty to fragment by CID. Also, in cases of mass overlap with coexisting species, selected fragmentation of a single component becomes difficult. We therefore applied the dual enzyme digestion method for improving MS/MS fragmentation information in a MALDI-Q-TOF instrument, hence improving the MS/MS method for protein identification.

5.3.6 MALDI MS/MS Fragmentation. Standard protein samples were digested with the trypsin and LAP digestion protocol and sampled at various times for MALDI analysis with a MDS Sciex API Q-TOF MS instrument. The detected fragments corresponding to the products of N-terminal digestion, along with their corresponding "parent" peptides, were selected for MS/MS fragmentation. The results obtained in these experiments are summarized below.

An example of an improvement in MS/MS fragmentation obtained using the trypsin and LAP digestion protocol is displayed in Figure 5.9. This figure depicts the MS/MS fragmentation spectra obtained from a standard solution digestion of 0.1 μg/μL BSA with trypsin and LAP. From the total MS scan (data not shown), a new peak at mass 1511.8 Da was observed which corresponds to N-terminal cleavage of a single Lysine residue from the parent peptide of mass 1639.9 Da. The MS/MS spectrum of the parent peptide is displayed in Figure 5.9A, and the MS/MS spectrum of the N-terminally processed fragment (mass 1511.8 Da) is shown in Figure 5.9B.

From Figure 5.9A, we can see that the MS/MS spectrum for the parent peptide of mass 1639.9 Da reveals only very weak signals corresponding to potential fragments of the parent ion. The MS-Product program was used to generate a list of all possible theoretical fragments of the parent peptide, however, no significant peaks in the MS/MS spectrum matched this theoretical list. In contrast to the results described above, the MS/MS spectrum obtained from the fragmentation of the N-terminally processed BSA peptide of mass 1511.8 Da revealed a number of significant peptide fragments, as shown in Figure 5.9B. When matched against a theoretical list of fragments generated with the MS-Product program, a number of detected peaks corresponded to fragments of the parent ion. The peaks corresponding to "y" ions and "b" ions are labeled in the figure. Still further peaks not labeled in the figure corresponded to either y or b ions with neutral loss of NH_3 or H_2O, or internal fragments of the parent peptide. These types of fragments are typical in a MALDI Q-TOF instrument. The observed fragments serve to confirm the

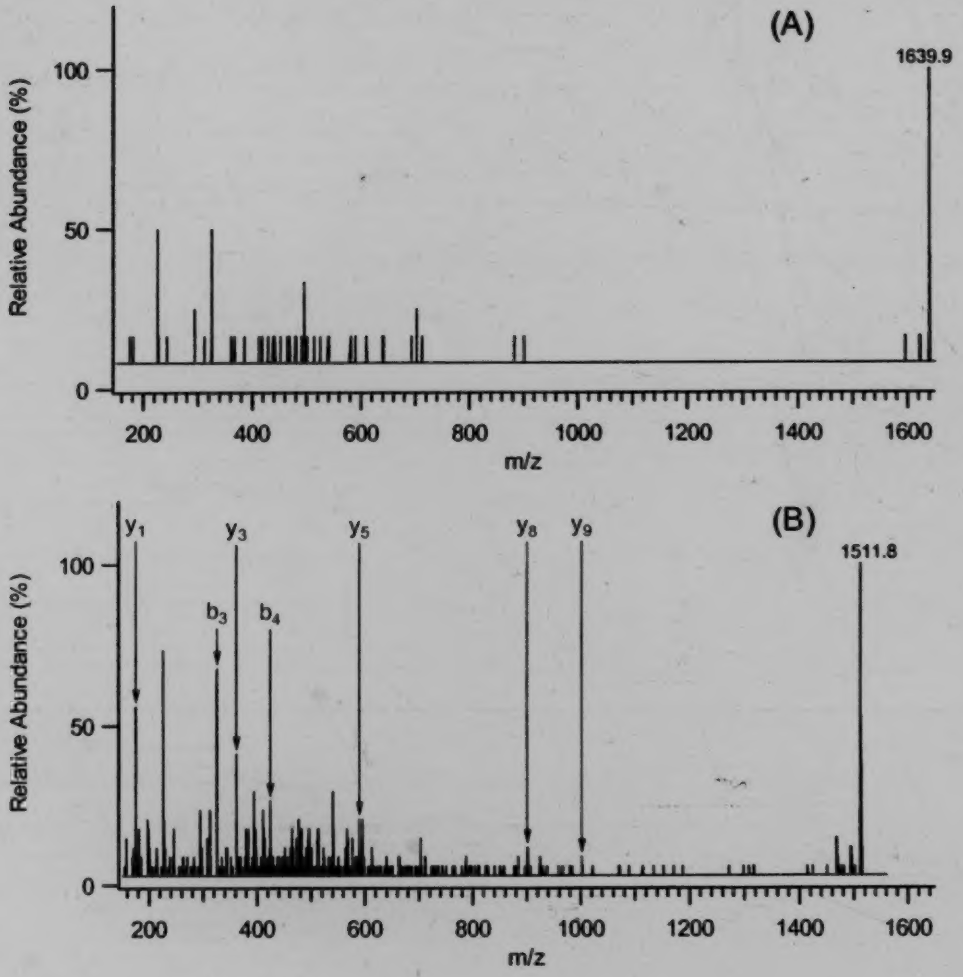

Figure 5.9 MALDI MS/MS spectra obtained from the fragmentation of the parent ion of mass (A) 1639.9 Da, and (B) 1511.8 Da on a Q-TOF instrument. The parent ions correspond to peptides from the solution digestion of 0.1 µg/µL BSA with trypsin and LAP. The peptide with mass 1511.8 Da results from N-terminal processing of the peptide at mass 1639.9 Da, resulting in cleavage of a single Lysine residue. The detected 'y' and 'b' type fragment ions are labeled in the figure.

sequence of the peak at 1511.8 Da. These results illustrate that N-terminal processing of peptides generated by tryptic digestion can be used to improve the quality of MS/MS spectra, and therefore provide increased information for protein sequence determination and protein identification.

Protein samples were also subjected to the in-gel digestion protocol following electrophoresis by one-dimensional SDS-PAGE. Figure 5.10 illustrates the sensitivity of this method in generating MS/MS fragmentation spectra for N-terminally cleaved peptides. In this experiment, a total of 100 ng of BSA was loaded on the gel. Following digestion with trypsin and LAP at various times, samples were analyzed on the MALDI Q-TOF instrument. The total MS scan for the BSA sample digested with trypsin alone is displayed in Figure 5.10A(i). The MS spectrum obtained from the analysis of a 10-minute LAP digestion is displayed in Figure 5.10A(ii). From these spectra, we can see that a total gel loading of 100 ng BSA was sufficient to produce several detectable BSA peptide fragments (labeled with an asterisk in the figure), as well as a new peak resulting from LAP cleavage of a Leucine residue from the BSA peak at mass 1479.8 Da. The N-terminally processed peak had a mass of 1367.7 Da, and is seen in the spectrum shown in Figure 5.10A(ii). These two peaks were then selected for MS/MS fragmentation, and the obtained spectra are displayed in Figures 5.10B (parent ion 1479.8 Da) and Figure 5.10 C (parent ion 1367.7 Da).

From the spectra displayed in Figure 5.10B-C, we can see that several peaks were detected that correspond to fragments of the parent ions. The y and b type ions have been labeled in the figure, although additional fragment ions were also observed. These additional fragments include several internal fragment ions, and fragment ions with neutral loss of NH_3 and H_2O.

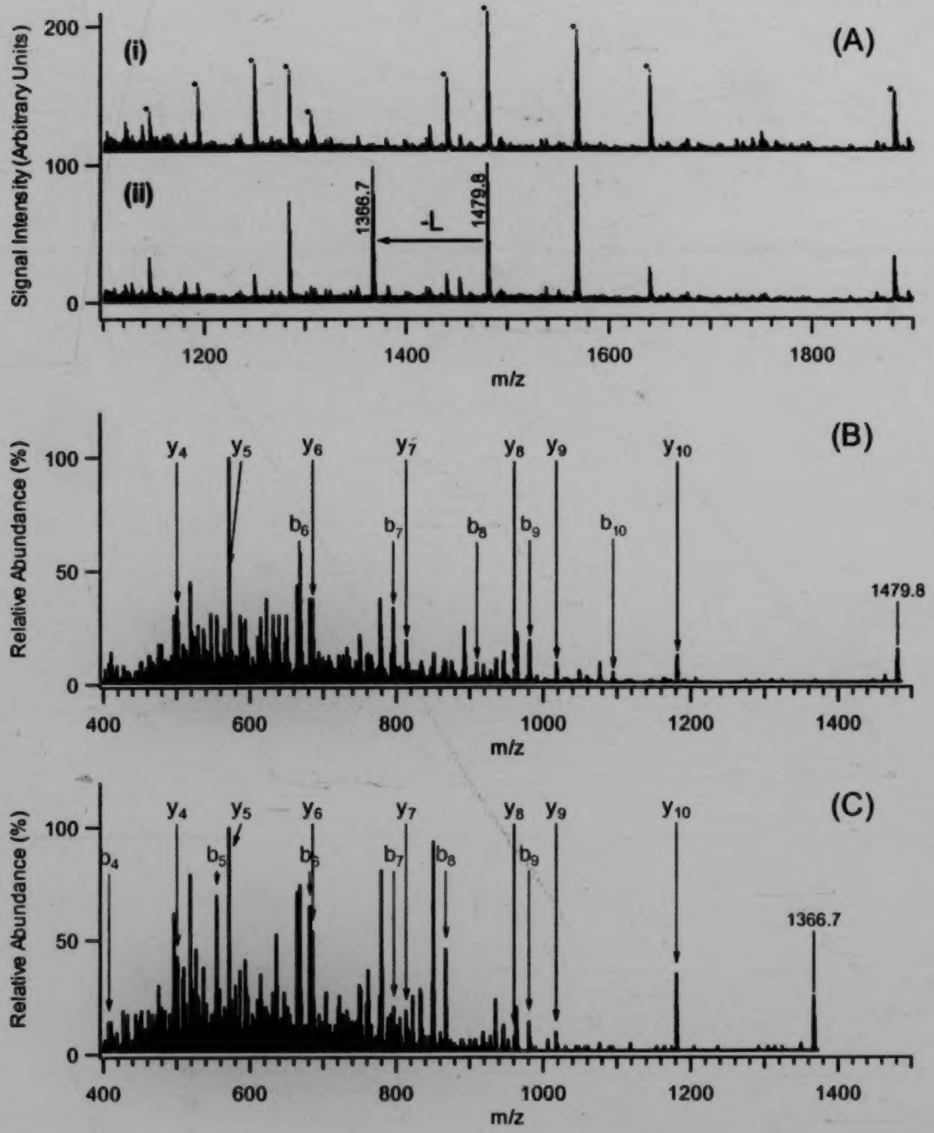

Figure 5.10 (A) MALDI MS spectra obtained from the analysis of 100 ng BSA on a q-TOF instrument. A(i) displays the spectrum from an in-gel tryptic digest of the sample and A(ii) is the spectrum obtained following LAP digestion for 10 minutes. (B) MALDI MS/MS spectrum obtained by fragmentation of the peptide with mass 1479.8 Da and (B) is the MS/MS spectrum from fragmentation of the peptide at mass 1366.7 Da. The b and y fragment ions are labeled in the figure.

As displayed in Figure 5.10, the fragmentation of both the "parent" BSA peptide and the N-terminally processed peptide revealed high quality MS/MS spectra, with several detectable fragment ions observed. The identical y ions were observed in each of these spectra. Unique b ions were observed in these spectra, resulting from N-terminal cleavage of the parent ion with LAP. The results of the MS/MS fragmentation of the N-terminally processed peptide with mass 1366.7 Da therefore serve to confirm the sequence of the peptide. This example serves to illustrate that the method is effective at generating fragmentation spectra from N-terminally processed peptides beginning with low amounts of protein (100 ng) isolated from polyacrylamide gels. We are currently applying this method for identification of unknown proteins.

5.4 Conclusions

The coupling of endoprotease with exoprotease digestion using trypsin followed by leucine aminopeptidase M was evaluated as a means of providing additional sequence information for protein identification. Given the proper digestion conditions, the enzyme LAP can be added directly to a trypsin-digested sample without the need for peptide separation. In doing so, several peptides in the mixture are digested by LAP, revealing amino acid sequence tags for these peptides. It is found that not all tryptic peptides will be digested by LAP, as it appears that this enzyme will preferentially digest certain amino acids over others (the non-polar amino acids). As well, the digestion efficiency appears to decrease over time and therefore only short amino acid sequence tags can be obtained for any given peptide. For these reasons, this methodology cannot be reliably used to sequence a specific peptide from a trypsin-digested sample. However, this work, along with work reported by others [19,20], clearly demonstrates that the additional information which is obtained from the generation of several small sequence tags can be very useful in providing additional sequence information for protein identification. The method was shown to be sensitive at the subpicomole level, and could provide additional sequence information in cases where nanospray MS/MS or post source decay failed to further sequence the peptides. This procedure was applied to HPLC separated *E. coli* fractions, as well as to a membrane protein from SDS PAGE, demonstrating that the method can be useful in identifying real world samples. In addition, the method was applied to improving MS/MS fragmentation spectra and served to confirm sequence information.

5.5 Literature Cited

(1) Schar, M.; Bornsen, K.; Gassmann, E. *Rapid Commun. Mass Spectrom.* **1991,** *5*, 319-326.

(2) Patterson, D. H.; Tarr, G.; Regnier, F. E.; Martin, S. A. *Anal. Chem.* **1995,** *67*, 3971-3978.

(3) Barlet-Jones, M.; Jeffery, W. A.; Hansen, H. F.; Pappin, D. J. C. *Rapid Commun. Mass Spectrom.* **1994,** *8*, 737-742.

(4) Vorm, O.; Roepstorff, P. *Biol. Mass Spectrom.* **1994,** *23*, 734-740.

(5) Chait, B. T.; Wang, R.; Beavis, R. C.; Kent, S. B. H. *Science* **1993,** *262*, 89-92.

(6) Cordwell, S. J.; Wilkins, M. R.; Cerpa-Poljak, A.; Gooley, A. A.; Duncan, M.; Williams, K. L.; Humphery-Smith, I. *Electrophoresis* **1995,** *16*, 438-443.

(7) Kaufmann, R.; Spengler, B.; Lutzenkirchen, F. *Rapid Commun. Mass Spectrom.* **1993,** *7*, 902-910.

(8) Katta, V.; Chow, D. T.; Rohde, M. F. *Anal. Chem.* **1998,** *70*, 4410-4416.

(9) Light-Wahl, K. J.; Loo, J. A.; Smith, R. D.; Witkowska, H. E.; Shackleton, C. H.; Wu, C. S. *Biol. Mass. Spectrom.* **1993,** *20,* 112-120.

(10) Scheler, C.; Lamer, S.; Pan, Z.; Li, X..P.; Salnikow, J.; Jungblut, P. *Electrophoresis*, **1998**, *19*, 918-927.

(11) Jensen, O. N.; Vorm, O.; Mann, M. *Electrophoresis* **1996,** *17*, 938-944.

(12) Reiber, D.; Grover, T. A.; Brown, R. S. *Anal. Chem.* **1998,** *70*, 673-683.

(13) Eng, J. K.; McCormack, A. L.; Yates III, J. R. *J. Am. Soc. for Mass Spectrom.* **1994,** *5*, 976-989.

(14) Mann, M.; Wilm, M. *Anal. Chem.* **1994,** *66*, 4390-4399.

(15) Medzihradsky, K. F.; Campbell, J. M.; Baldwin, M. A.; Falick, A. M.; Juhasz, P.; Vestal, M. L.; Burlingame, A. L. *Anal. Chem.* **2000,** *72*, 552-558.

(16) Loboda, A. V.; Krutchinsky, A. N.; Bromirski, M., Ens, W.; Standing, K. G. *Rapid Comm. Mass. Spectrom.* **2000,** *14*, 1047-1057.

(17) Bonetto, V.; Bergman, A.-C.; Jornvall, H.; Sillard, R. *Anal. Chem.* **1997,** *69*, 1315-1319.

(18) Schriemer, D.C.; Yalcin, T.; Li, L. *Anal. Chem.* **1998,** *70*, 1569-1575.

(19) Korostensky, C.; Staudenmann, W.; Dainese, P.; Hoving, S.; Gonnet, G.; James, P. *Electrophoresis* **1998,** *19*, 1933-1940.

(20) Owens, D. R.; Bothner, B.; Phung, Q.; Harris, K.; Siuzdak, G. *Bioorganic & Medicinal Chemistry* **1998,** *6*, 1547-1554.

(21) Maclean, G. D.; Seehafer, J.; Shaw, A. R. E.; Kieran, M. W.; Longenecker, B. M. *J. National Cancer Institute* **1982,** *69*, 357-363.

(22) Wang, Z. P.; Russon, L.; Li, L.; Roser, D. C.; Long, S. R. *Rapid Commun. Mass Spectrom.* **1998,** *12*, 456-464

(23) Dai, Y.; Randy M. Whittal, R.M; Li, L. *Anal. Chem.* **1999,** *71*, 1087 –1091.

(24) Whittal, R. M.; Li, L. *Anal. Chem.* **1995,** *67*, 1950-1954.

(25) Burrel, M. M., Ed. Methods in Molecular Biology, Vol. 16: Enzymes of Molecular Biology, **1993,** Humana Press Inc.: Totowa, NJ.

(26) Lorand, L., Ed. Methods in Molecular Biology, Vol. 45: Proteolytic Enzymes, Part B. **1976,** Academic Press: New York.

Chapter 6

Enrichment of Phosphorylated Proteins From a Mixture with Immobilized Metal Affinity Chromatography

6.1 Introduction

The previous chapters presented methods related to the characterization of the primary structure of proteins with mass spectrometry. However, the applicability of mass spectrometry for protein characterization can extend beyond amino acid sequence determination. MS-based techniques can also be used to determine post-translational modifications of a protein [1-3]. Of the well over 200 known protein modifications, phosphorylation is of particular biological relevance [4]. As it is a reversible process, protein phosphorylation is a key regulator of protein function and activity in biological processes [4-6]. The regulation of protein phosphorylation/ dephosphorylation is dictated by two families of counteracting enzymes, namely the protein kinases and phosphatases [6].

The most common form of phosphorylation involves the phosphate ester linkage of a phosphate group with the hydroxyl group on the side chain of serine, threonine, and tyrosine [5]. The phosphorylated forms of these three amino acids are shown in Figure 6.1. A molecular weight shift of +80 Da is associated with the addition of a phosphate group to these amino acids, and thus MS based methods can be used to detect, as well as localize this modification on the polypeptide chain of a protein.

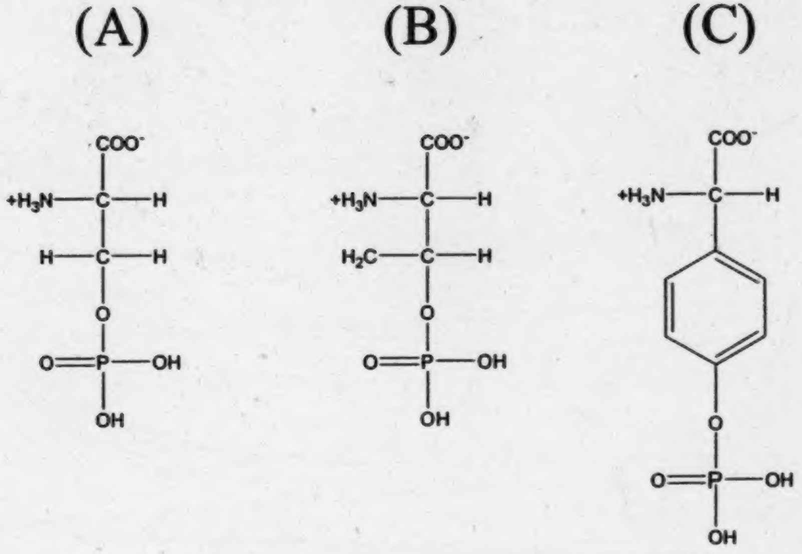

Figure 6.1 Chemical structures of (A) phosphoserine, (B) phosphothreonine, and (C) phosphotyrosine.

A major difficulty associated with the determination of protein phosphorylation by mass spectrometry is related to the low stoichiometric ratio in which proteins are phosphorylated *in vivo* [5]. Thus, a need for phosphoprotein enrichment is imposed prior to MS analysis in order to achieve sufficient quantity of sample to meet the detection limits associated with MS techniques. Sample enrichment involves the separation of the phosphorylated from the non-phosphorylated protein components present in a mixture. Of the separation methods available, immobilized metal affinity chromatography (IMAC), a form of ligand-exchange chromatography, has been explored for selective enrichment of phosphoproteins prior to MS analysis [7,8].

IMAC is based on the selective affinity of metal ions for various chemical groups [9,10]. The metal ion is chelated to a chromatographic support through a ligand attachment, such as iminodiacetic acid or tris(carboxymethyl)ethylenediamine [11-13]. It has been found that Fe^{3+} and Ga^{3+} have affinity for the electron-donor atoms, such as oxygen in a phosphate group [7,8,14]. Coordinative binding of the oxygen with the unoccupied coordination sites of the chelated metal ions result imparts an affinity of phosphorylated molecules with the metal-charged surface. Elution of the bound molecules can be accomplished by competitive displacement of the coordinated phosphate groups [10]. This is most simply accomplished by elution of the IMAC support with a solution containing sodium phosphate [14]. Elution of phosphopeptides from an IMAC support has also been accomplished by raising the pH of the solution to highly basic values with a 2% solution of NH_4OH [15]. Although the mechanism of this elution has not been described, it is presumed that hydroxide ions present in the basic solution will act to competitively displace the bound phosphopeptides from the metal-charged support.[9]

Thus, IMAC supports with these metal ions have been applied to the separation or enrichment of phosphoproteins. Recent efforts with IMAC enrichment have been focused on the selective capture of phosphopeptides, rather than intact phosphoproteins. However, purification of the intact phosphoprotein has some benefits in that it allows for increased sample loading on gels, improved digestion and detection efficiencies, and a decrease in background from non-phosphorylated protein sample.

---

[9] The mechanism to describe the elution of phosphopeptides from an IMAC support at high pH has been proposed by Dr. Zhengping Wang, Department of Chemistry, University of Alberta.

In this chapter, the use of IMAC for the enrichment of phosphopeptides and phosphoproteins is explored. The selectivity of a Fe^{3+}-immobilized sepharose support is demonstrated using a sample of cow's milk. An initial study on the applicability of the IMAC technique for facile enrichment of phosphopteins from a mixture is performed using the standard, α-casein, as a model phosphoprotein.

6.2 Experimental

6.2.1 Materials and Reagents. Cytochrome c, as well as trypsin (from bovine pancreas, TPCK treated to reduce chymotrypsin activity; dialyzed, lyophilized) and alkaline phosphatase (EC number 3.1.3.1) were from Sigma Aldrich Canada (Oakville, ON). The α-casein used in these experiments was from Fluka. HiTrap Chelating HP sepharose beads were from Amersham Pharmacia Biotech (Uppsala, Sweden), and were stored in a 20% ethanol/water solution at 4°C prior to use. The 1-200 µL Micro capillary Tips (gel loader tips) used to form the columns were purchased from Rose Scientific (Edmonton, AB). Analytical grade acetone, methanol, acetonitrile, acetic acid, ammonia, and trifluoroacetic acid (TFA) were purchased from Caledon Laboratories (Edmonton, AB). Water used in the experiments was from a NANOpure water system (Barnstead/Thermolyne). Skimmed cow's milk was obtained from a local grocery store. The chemicals required for casting of SDS polyacrylamide gels were from Bio-Rad. All other chemicals were from Sigma.

6.2.2 IMAC Column Preparation. The HiTrap Chelating HP sepharose beads were charged with Fe^{3+} metal ions prior to loading the beads into microcolumns. Approximately 200 μL of the sepharose beads were twice washed in a vial with 1 mL of water, followed by two washes with 1 mL of 0.1 M acetic acid. Next a 1-mL solution of 100-mM $FeCl_3$ in 0.1 M acetic acid was added to the beads, and the beads were vortexed for several minutes before centrifugation and removal of the supernatant. The beads were again washed with five changes of 0.1 M acetic acid to remove the non-chelated Fe^{3+}, and finally suspended in 1 mL of 0.1 M acetic acid. The IMAC microcolumns were formed by packing the sepharose beads in gel loader tips, whose ends had been pinched with pliers in order to retain the beads. Individual IMAC columns were packed with ~2 μL of sepharose beads.

6.2.3 Sample Loading, Washing, and Elution from IMAC Columns. The protein samples, acidified to a pH of ~3.5 with acetic acid, were loaded on individual IMAC columns by flowing the solution through the column at an approximate flow rate of 5 μL/min using pressurized nitrogen. Following sample loading, the column bed was washed with 50 to 100 μL of 0.1 M acetic acid, or acetic acid containing various percentages of acetonitrile. Samples were eluted from the IMAC column with 20 μL aliquots of NH_4OH prepared at various concentrations (between 0.01% and 3% of saturated) to produce basic solutions of increasing pH values. The IMAC columns were also eluted with a solution of 50% acetonitrile in 0.1% aqueous TFA.

6.2.4 Enzyme Digestions. Digestion of protein samples with trypsin was performed using standard digestion protocols. For solution digests, samples were buffered in 0.1 M NH_4HCO_3 with the addition of 1 M NH_4HCO_3 to the sample. The standards were digested with trypsin in a 10:1 mass ratio of protein to trypsin for 1 hour at 37°C. For the milk sample, 1 mL of 50:1 diluted milk was digested with 5 μg trypsin for 1 hour, omitting sample reduction and alkylation. Samples were also digested directly following adsorption to IMAC resin in a vial. To the IMAC resin containing adsorbed protein, 10 μL of a buffered trypsin solution (0.2 μg trypsin) was added and the sample was digested for 1.5 hours at 37°C before sampling the bead slurry for MALDI analysis.

Protein dephosphorylation was performed using the enzyme alkaline phosphatase. The enzyme stock solution was diluted by a factor of 50:1 in 0.1 M NH_4HCO_3. A 5-μL aliquot of the diluted enzyme was added to 100 μL of the protein solution containing the phosphorylated sample components. The sample was digested at 37°C for 1 hour.

6.2.5 Gel Separation, Staining and Imaging. Protein samples were separated by 1-dimensional SDS page on a 15% Tris-Tricine polyacrylamide gel cast in-house. Prior to gel loading, the protein samples were reduced by heating the sample for 5 minutes at 95°C in a solution containing DTT. The samples were separated with a constant current setting of 20 mA per gel with a Bio-Rad mini-PROTEAN 3 Electrophoresis System. The gels were stained with Bio-Safe Coomassie blue (Bio-Rad, Mississauga, ON) and imaged with a gel scanner.

6.2.6 MALDI Sample Preparation. Recrystallized α-cyano-4-hydroxycinnamic acid (HCCA) was used as matrix in these experiments. The samples were deposited using a two-layer method that has been described in detail elsewhere [16]. The sample was mixed with second layer matrix solution in a 1:10 (v/v) ratio. The second layer matrix solution was either prepared as a saturated solution in 40% methanol/0.1% aqueous TFA, or in a 1:2:3 mixture of formic acid/isopropanol/water. This second solvent system was exclusively used in the analysis of the undigested milk sample, which was also mixed with matrix in a 1:10 (v/v) ratio.

6.2.7 Instrumentation. Mass spectra were collected on a time-lag focusing MALDI time-of-flight mass spectrometer, which was constructed at the University of Alberta and has been described in detail elsewhere [17]. The MALDI spectrum for the undigested milk sample was recorded using a Bruker Reflex III MALDI time-of-flight system (Bremen/Leipzig, Germany). MALDI MS/MS fragmentation experiments were performed on a MDS Sciex API Q-Star quadrupole-time-of-flight mass spectrometer equipped with a MALDI source (Toronto, ON). A pulsed nitrogen laser, operating at 337 nm, was used to generate the MALDI ions. The reported mass spectra are the result of signal averaging of between 50 to 100 collected spectra. Data were processed using the IGOR Pro software package (Wavemetrics Inc., Lake Oswego, OR).

6.3 Results and Discussion

6.3.1 Phosphopeptide Enrichment from Cow's Milk. The enrichment of phosphopeptides can be accomplished by exploiting the selectivity of immobilized metal affinity chromatography using Fe^{3+} or Ga^{3+}. These metal ions have proven to possess an affinity for oxygen on the phosphate group(s) of a phosphopeptide. Here, an IMAC micro-column, packed with Fe^{3+}-charged sepharose beads, is used to enrich phosphorylated peptides from a sample mixture. The selectivity of these self-packed IMAC columns is first demonstrated using a 'real world' sample of cow's milk. Although several different types of proteins are present in cow's milk, the main protein components include α-lactalbumin, β-lactoglobulin, and various caseins (α_{s1}, α_{s2}, and β-casein). Of these proteins, α_{s1}-casein contains 7 phosphorylated serine residues, α_{s2}-casein contains 10 phosphoserines, while β-casein is phosphorylated at 4 serine residues.

Figure 6.2A displays a MALDI spectrum recorded from the MS analysis of a sample of cow's milk that had been diluted in water by a factor of 1:50 (v/v). The main protein components of the sample exhibit intense signals in this MALDI spectrum, and the singly charged molecular ions of these species have been labeled in the figure. Along with signals from the doubly protonated species of these compounds, several additional signals are also observed in this spectrum. Many of these peaks can be attributed to fragments of α_{s1} and β-casein, resulting from plasmin digestion of these proteins in the milk sample [18]. The diluted milk was then subjected to digestion with trypsin, and the MALDI spectrum obtained by analyzing a portion of this digest is displayed in Figure 6.2B. Most of the observed peaks in this spectrum have m/z values that correspond to theoretical peptide fragments of α-lactalbumin, β-lactoglobulin, or the various caseins. Although a large number of intense peaks are visible in this spectrum, it is expected that many additional components have not been detected from this direct MALDI analysis of the digest, as a result of signal suppression. The digested milk sample was then subjected to IMAC purification. The digest was first acidified to a pH of ~3.5 using 10% acetic acid. The sample was then passed through an IMAC column, formed by packing a gel-loader tip with 2.5 μL Fe^{3+}-IMAC resin, at an approximate flow rate of 5 μL/min. The IMAC column was then washed with 100 μL of 0.1 M acetic acid containing 10% acetonitrile. The wash solution was collected in 20-μL fractions and analyzed by MALDI-MS. The results revealed strong signals from the first fraction, however no detectable signals were observed in the final wash (results not shown).

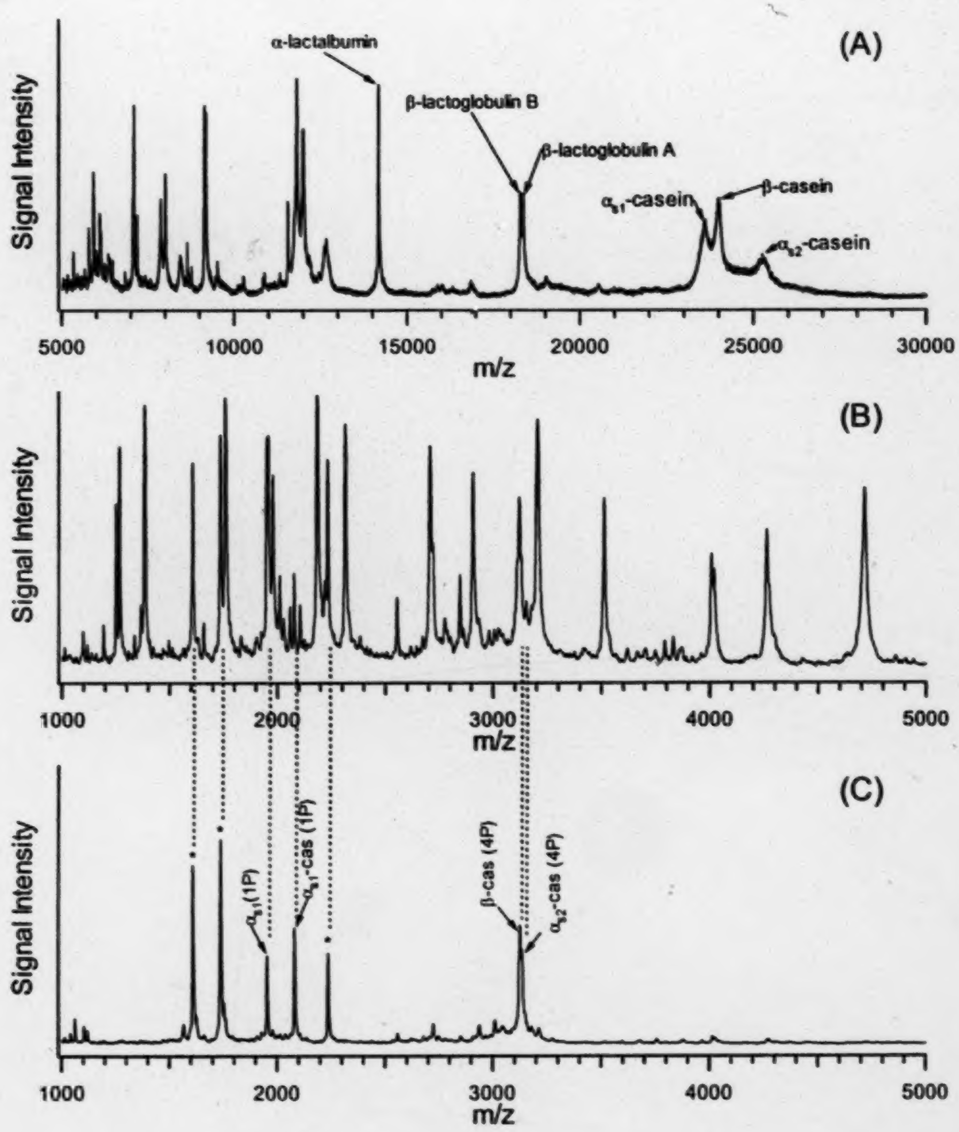

Figure 6.2 MALDI spectra recorded from the analysis of a sample of 1:50 diluted cow's milk. (A) is the spectrum from the undigested sample, (B) is for the sample of milk that had been digested with trypsin, and (C) was obtained following IMAC enrichment of the tryptic digest. The main protein components are labeled in (A), while the peaks corresponding to phosphopeptides are labeled in (C). The peaks labeled with an asterisk in (C) represent non-phosphorylated peptides.

The IMAC column was subsequently eluted with 20 μL of a 2% saturated NH_3 solution in water (pH ~12). The basic extraction was subjected to MALDI analysis, and the resulting spectrum is displayed in Figure 6.2C.

The spectrum in Figure 6.2C exhibits seven peaks of strong signal intensities. As seen in Figure 6.2, these peaks were also observed in the MALDI spectrum obtained from the raw digest, however these seven peaks represent only a small subset of the total peptides present in the raw tryptic digest. Based on mass assignment of the peaks observed after IMAC purification, four peaks were attributed to phosphorylated peptide fragments of α_{s1}, α_{s2}, and β-casein (labeled in Figure 6.2C). These results were confirmed by alkaline phosphatase digestion of the raw tryptic digest, which revealed a mass shift for each of these peptides corresponding to cleavage of the phosphate residues (results not shown). It was concluded that the remaining three peaks observed in the MALDI spectrum following IMAC enrichment were non-phosphorylated peptides. The mass of these peaks did not correspond to any casein phosphopeptides. Also, the alkaline phosphatase digestion did not produce a mass shift for these peptides.

An attempt was made to assign sequence identity for the three unknown peaks in the MALDI spectrum of Figure 6.2C. The mass of these peaks did not match any peptides resulting from theoretical tryptic digestion of the main protein components of milk. However, MALDI MS/MS analysis of one of these unknown peptides, having a mass of 1607.7 Da, revealed a high quality fragmentation spectrum. The fragmentation ions were imported into the MASCOT search program; however, a potential match of the peptide sequence could not be found. Therefore, a *de novo* approach was taken in order to determine the amino acid sequence of the peptide. By manual interpretation of the obtained MS/MS spectrum, a nearly complete amino acid sequence for this unknown peptide could be determined, based on a series of b-ions observed in the MS/MS spectrum. The obtained sequence is given below in Chart 6.1.

Chart 6.1 Amino acid sequence of the peptide with an observed mass of 1607.7 Da, as determined by *de novo* sequencing.

X-X-X[a]-(I/L[b])-E-E-(I/L)-N-V-P-G-E-I-V-E [10]

[a] The N-terminal amino acids could not be determined from *de novo* sequencing. However, a mass deficiency of 267 Da must constrict these residues to PGI or PAV, assuming no modifications (order unknown).

[b] The amino acid residue at this position is assignment to either leucine (L) or isoleucine (I).

From the amino acid sequence of the peptide given in Chart 6.1, one can see that this peptide is highly acidic, having 4 glutamic acid (E) residues. The high number of acid residues on this peptide provides a likely explanation as to why this particular peptide would be enriched in the IMAC purification. It is known that Fe^{3+} ions have strong affinity for acidic groups, such as the carboxyl group on the side chain of glutamic acid. As such, acidic peptides will bind to the immobilized metal of the Fe^{3+}-sepharose resin, and are therefore observed following the IMAC purification procedure. Although

---

[10] MASCOT sequence searching of a portion of this sequence matched a thioredoxin protein from *Lactobacillus bulgaricus* (TrEMPLE entry Q93FC9). *L. bulgaricus* is a type of gram-positive bacteria that is involved in the fermentation of milk, as the bacteria converts sugar to lactic acid.

MALDI MS/MS could not determine the sequence of the remaining two peptides, it is suspected that these peptides might also contain a high degree of acidic residues, which would explain their affinity for the Fe^{3+}-IMAC column.

6.3.2 Enrichment of Phosphoproteins by IMAC Purification. The analysis of phosphoproteins by mass spectrometry is most often accomplished using the approach described in the previous section, which involved protein digestion prior to phosphopeptide enrichment with IMAC. However, the enrichment of intact phosphoproteins, prior to tryptic digestion, can offer several potential advantages. First, by removing the non-phosphorylated components from the system, those species that would add to background interference in the MS analysis are eliminated. Second, by enriching the phosphoprotein prior to digestion, one can improve digestion kinetics and digestion efficiency. Third, reducing the background components can offer an advantage in SDS-PAGE electrophoresis, in cases where the phosphoprotein makes up a low percentage of the total protein content in the sample. Phosphoprotein enrichment will therefore allow for increased loading of the components of interest on the gel. Also, possible co-migration of phosphoproteins with non-phosphorylated protein can be reduced.

The selectivity of Fe^{3+}-charged sepharose beads for an intact phosphoprotein was explored using a mixture of undigested α-casein and cytochrome c. In this procedure, the mixture was loaded on the sepharose beads in a vial, and following a washing step, the protein component that remained bound to the IMAC resin was digested using a bead digestion procedure, in which trypsin is directly added to the sample while adsorbed to the beads.

Figure 6.3A displays a MALDI spectrum obtained from digestion of a 100 μL solution of 1 μM α-casein, following adsorption to 2 μL of Fe^{3+}-sepharose beads. In this experiment, following loading of the α-casein, the beads were washed with 0.1 M acetic acid. As can be seen from Figure 6.3A, the spectrum reveals that the α-casein has effectively adsorbed to the IMAC resin, producing a very complete peptide mass map for the protein following digestion. The selectivity of the IMAC beads for adsorption of phosphoprotein was explored with a 100 μL solution of 1 μM α-casein and 1 μM cytochrome c, which was mixed with 2 μL Fe^{3+}-sepharose beads in a vial. Again, the beads were washed with 0.1 M acetic acid prior to digestion with trypsin. The obtained MALDI spectrum is displayed in Figure 6.3B. The results of this experiment revealed that the IMAC resin did not exclusively adsorb the phosphoprotein. As seen in the MALDI spectrum, several peptide fragments of cytochrome c were observed along with the α-casein peptide fragments. This therefore indicates that cytochrome c was also adsorbed to the Fe^{3+}-sepharose beads. This result might be expected as one considers the composition of sepharose support for the IMAC resin. The sepharose support can act as an effective surface for adsorption of hydrophilic proteins such as cytochrome c. As a result, attempts at reducing non-specific adsorption of cytochrome c on the IMAC resin, or in washing the cytochrome c from the support were unsuccessful. This includes loading and washing the column with solutions containing high percentages of acetonitrile. However, when the IMAC resin was washed with a solution of NH_4OH (pH ~12) prior to tryptic digestion of the bound protein components, an interesting result was obtained. The MALDI spectrum from this digestion is shown in Figure 6.3C. As seen in the spectrum, when the IMAC beads were washed with a basic solution prior to digestion,

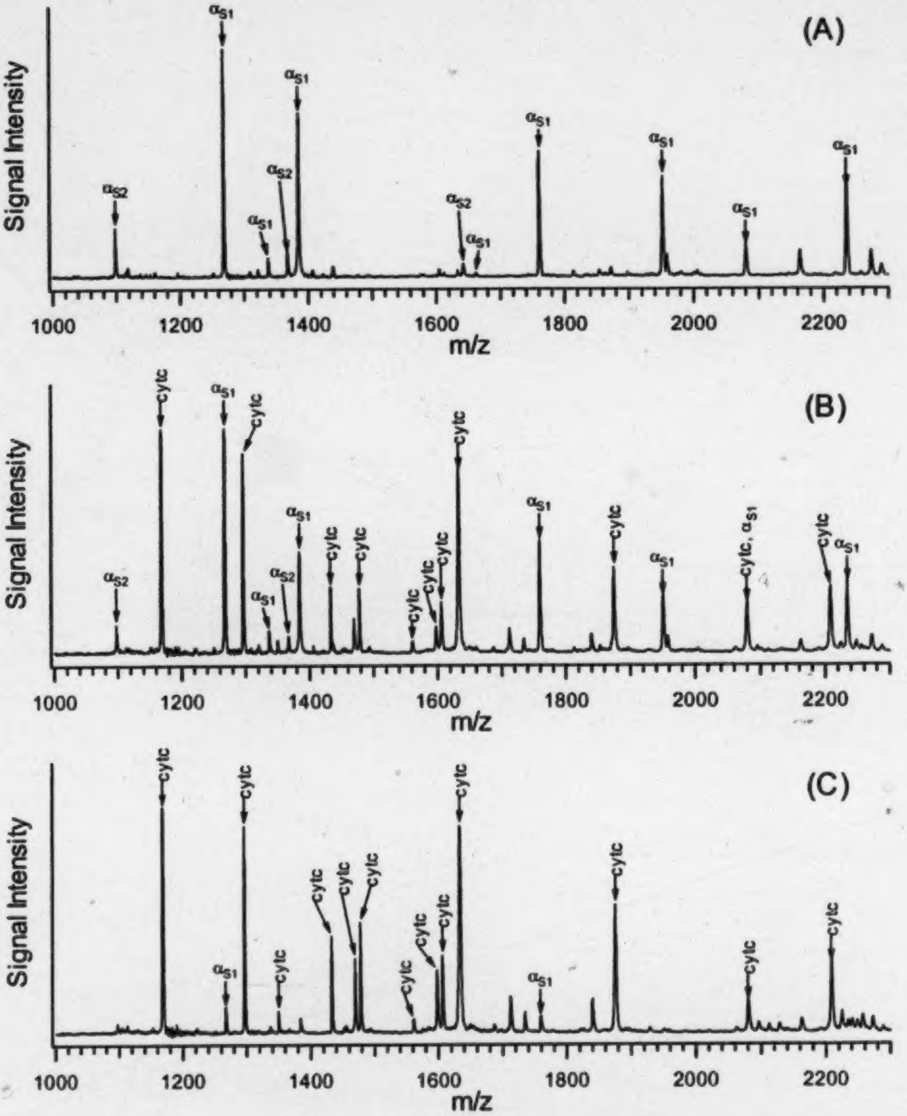

Figure 6.3 MALDI spectra obtained from the analysis of tryptic digests of protein samples following adsorption and washing on IMAC beads. (A) was obtained by loading α-casein on IMAC beads, with an acetic acid wash (pH 3.5) (B) was from loading α-casein and cytochrome c, with acetic acid wash (pH 3.5), and (C) was from loading α-casein and cytochrome c, with an NH_4OH wash (pH 12) prior to digestion. Peptide fragments of cytochrome c and of α-casein (α_{S1} or α_{S2}) are labeled 'cytc' or 'α_{S1}' and "α_{S2}' respectively.

only peptide fragments of cytochrome c could be detected. From this, it can be concluded that the basic wash effectively removed the α-casein from the Fe^{3+}-sepharose beads, while a part, or all of the cytochrome c remained bound to the beads.

The results described above are further explored by directly analyzing the solutions used to wash the IMAC beads following sample loading of α-casein and cytochrome c. An IMAC column was therefore charged using a solution of 1 μg α-casein and 1 μg cytochrome c, prepared in 0.1 M acetic acid. Approximately 2 μL of IMAC resin was used to form the column. Following a washing step with 0.1 M acetic acid, and water, the column was successively extracted with a basic NH_4OH solution (pH ~9.5), and then with a solution of 50% acetonitrile in 0.1% TFA. The basic extract, as well as the organic solvent extract, were loaded into individual lanes of an SDS polyacrylamide gel and subjected to 1D gel electrophoresis. For comparison, a mixture of 1 μg cytochrome c and 1 μg α-casein was also loaded on the gel in a separate lane. Following electrophoresis, the gel was stained with coomassie blue and imaged with a gel scanner. The resulting image is displayed in Figure 6.4.

Lane I of Figure 6.4 contains the basic elution from the IMAC column loaded with α-casein and cytochrome c. Lane II contains the 50% acetonitrile/ 0.1% TFA extract of the IMAC column, and Lane III is the standard mixture of 1 μg α-casein and 1 μg cytochrome c. From comparison to Lane III, a single band corresponding to α-casein is clearly visible in the gel image in Lane I. The intensity of this band is somewhat lower than the 1 μg standard of α-casein. No band for cytochrome c was observed in this lane. This therefore indicates that the intact α-casein was adsorbed, and subsequently eluted from the IMAC column through a change in pH. Less than 100% recovery of the

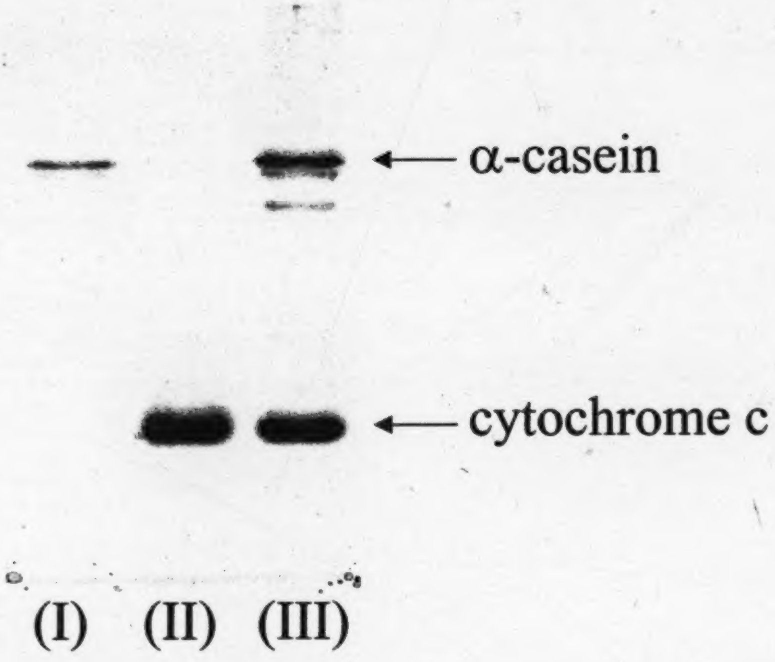

Figure 6.4 Gel image obtained from electrophoresis of (I) the basic extract (pH ~9.5) of an IMAC column loaded with α-casein and cytochrome c (II) a 50% acetonitrile/0.1% TFA extraction of the same column following the first extraction and (III) a standard consisting of 1 μg α-casein and 1 μg cytochrome c.

α-casein from the IMAC purification is attributed to partial loss of the sample during the extensive washing of the IMAC column. Lane II of the gel from Figure 6.4 displays a strong band for cytochrome c, while no band is detected corresponding to α-casein in this elution. The intensity of the cytochrome c band is comparable to the 1 μg cytochrome c standard in Lane III, indicating that most of the cytochrome c had adsorbed to the IMAC column. Also, the cytochrome c was retained on the column during both the washing steps, as well as the basic extraction, until it was finally eluted from the resin with 50% acetonitrile/ 0.1% TFA. These results clearly demonstrate that while the IMAC resin does not exhibit any selectivity for capture of α-casein and cytochrome c, the enrichment capabilities of the IMAC column can be exploited through selective elution of α-casein from the column using a basic extraction with NH_4OH (pH ~9.5).

As mentioned in section 6.1, the binding and elution of α-casein from the IMAC column relies on a change in pH to alter the affinity of the protein for the IMAC resin. At low pH, α-casein binds to the IMAC resin, likely through interactions between immobilized iron (III) ions and the many exposed phosphate groups of the α-casein. As the pH of the solution is raised, the increasing concentration of free hydroxide ions will competitively replace the phosphate groups bound to the Fe^{3+}, resulting in the elution of α-casein from the resin at high pH. As the cytochrome c is suspected to adsorb directly to the sepharose support, as opposed to the immobilized Fe^{3+} of the IMAC resin, the increase in pH does not significantly disrupt the binding of this protein to the support. Since the release of α-casein from the IMAC resin is determined by a change in pH, the effect of pH on the binding and elution properties of the IMAC resin was further explored.

Four individual IMAC columns were charged with a mixture of 10 μg each cytochrome c and α-casein, and the columns were subsequently washed with acetic acid (pH ~3.5) and then with water (pH 5.5). The individual columns were then extracted with aqueous solutions prepared at different pH values. The extracts were then buffered with 0.1 M NH_4HCO_3, adjusting the pH to ~8, and digested in solution with trypsin. The

MALDI spectra obtained from the analysis of the resulting digests are displayed in Figure 6.5. Figure 6.5A is the spectrum from digest of the extract with pH 7.5, Figure 6.5B is the spectrum from an extract at pH 9, and Figure 6.5C is from an extract at pH 12. Also, an IMAC column was extracted with 50% acetonitrile/ 0.1% TFA and the MALDI spectrum from the digest of this extract is shown in Figure 6.5D.

From Figure 6.5A, we can see that the pH 7.5 extraction almost exclusively resulted in the detection of α-casein from the digestion of the extract. Only 3 weakly observed peaks from cytochrome c fragments were observed in this spectrum, indicating that very little cytochrome c has eluted at this pH. At a higher pH of 9 (Figure 6.5B), we again see a high degree of α-casein peaks in the spectrum. However, an increase in both the number, and signal intensity of cytochrome c peptide fragment peaks was observed in the MALDI analysis of the fraction. This indicates that a higher degree of cytochrome c has been extracted from the IMAC column. Recall from Figure 6.4 that a band for cytochrome c was not observed in the basic extraction at pH 9.5. In this case, a higher amount of cytochrome c was loaded on the column. Also, the MALDI detection method is known to be more sensitive than the limits of coomassie staining. At a pH of 12, the MALDI spectrum reveals even more cytochrome c peptide fragments, indicating that a higher degree of cytochrome c is eluted as the pH is further increased from 9 to 12. The 50% acetonitrile extraction of the IMAC column would result in total elution of both proteins from the column. As observed in Figure 6.5D, a significant number of cytochrome c peaks are observed, along with fragments of α-casein. The relative intensity of the cytochrome c peaks is highest for the organic elution. Although MALDI

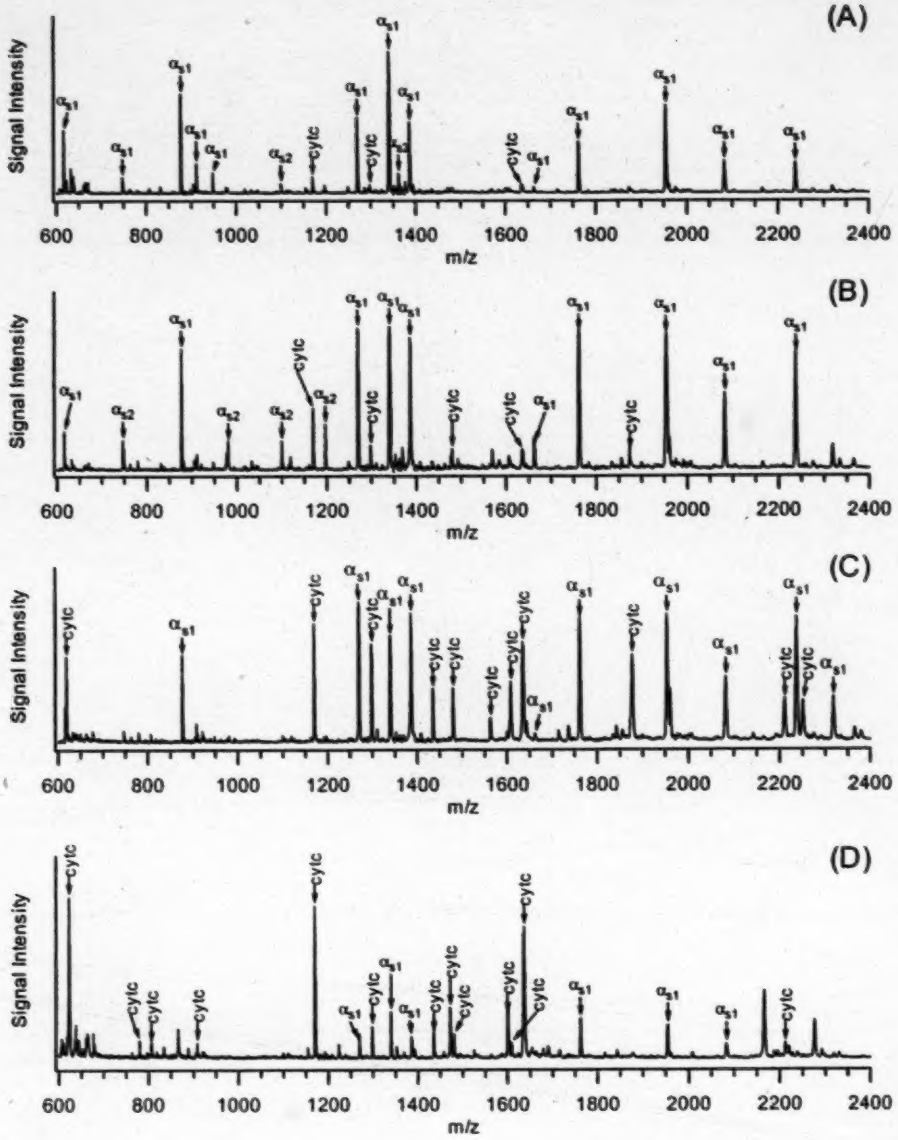

Figure 6.5 MALDI spectra obtained from tryptic digests of a mixture of α-casein and cytochrome c following an IMAC purification procedure, with extraction from the IMAC column with various solutions. (A) The IMAC column was extracted with an aqueous solution at pH 7.5, (B) at pH 9, (C) at pH 12, and (D) extracted with a solution of 50% acetonitrile/0.1% aqueous TFA. The peaks are labeled as in Figure 6.4.

results are not strictly quantitative, these results indicate that while a change in pH to basic values will result in the elution of α-casein from the IMAC column, a further increase in pH will elute an increasing amount of cytochrome c from the column as well. Therefore, in order to exploit the full selectivity of the IMAC resin for enrichment of α-casein, the pH of the eluting solution must be carefully controlled.

6.4 Conclusions

Immobilized metal affinity chromatography using a Fe^{3+}-charged sepharose support was demonstrated to show selectivity enrichment of phosphopeptides from a mixture. It is noted that the results described in this chapter do not preclude the role of the sepharose support in the binding and elution of proteins at varied pH. Further experiments should therefore be conducted to determine the specific role of immobilized metal in determining the affinity of the Fe^{3+}-IMAC support. The binding of acidic peptides to this resin also results in the enrichment of such components using the IMAC purification procedure. The IMAC enrichment procedure was also applied to the enrichment of intact α-casein from a mixture of α-casein and cytochrome c. In this case, it was seen that although cytochrome c binds to the resin, the phosphoprotein could be selectively eluted from the support with a basic extraction. These results therefore demonstrate the possibility of using IMAC supports for enrichment of high molecular weight phosphoproteins from a mixture. Future work is focused on the application of the IMAC method for enrichment of other phosphoproteins, including the selective enrichment of phosphoproteins from their corresponding non-phosphorylated species after an *in vitro* phosphorylation experiment on a protein of interest.

6.4 Literature Cited

(1) Aebersold, R.; Goodlett, D. R. *Chem. Rev.* **2001**, *101*, 269-295.

(2) Demirev, P. A.; Lin, J. S.; Pineda, F. J.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 4566-4573.

(3) Wilkins, M. R.; Gasteiger, E.; Gooley, A. A.; Herbert, B. R.; Molloy, M. P.; Binz, P. A.; Ou, K.; Sanchez, J. C.; Bairoch, A.; Williams, K. L.; Hochstrasser, D. F. *J. Mol. Biol.* **1999**, *289*, 645-657.

(4) Larsen, M.; Roepstorff, P. *Fresenius J. Anal. Chem.* **2000**, *366*, 677-690.

(5) Kaufmann, H.; Bailey, J. E.; Fussenegger, M. *Proteomics* **2001**, *1*, 194-199.

(6) Johnson, L. N.; Lewis, R. J. *Chem. Rev.* **2001**, *101*, 2209-2242.

(7) Andersson, L.; Porath, J. *Anal. Biochem.* **1986**, *154*, 250-254.

(8) Muszynska, G.; Andersson, L.; Porath, J. *Biochemistry* **1986**, *25*, 6850-6853.

(9) Chaga, G. *J. Biochem. Biophys. Methods* **2001**, *49*, 313-334.

(10) Gaberc-Porekar, V.; Menart, V. *J. Biochem. Biophys. Methods* **2001**, *49*, 335-360.

(11) Rassi, Z. E.; Horvath, C. *J. Chromatography* **1986**, *359*, 241-252.

(12) Porath, J.; Olin, B. *Biochemistry* **1983**, *22*, 1621-1630.

(13) Porath, J.; Olin, B.; Granstrand, B. *Archives Biochem. Biophys.* **1983**, *225*, 543-547.

(14) Posewitz, M. C.; Tempst, P. *Anal. Chem.* **1999**, *71*, 2883-2892.

(15) Brucato, C. L.; Asara, J. M.; Pierce, K. A.; Schnitzler, A.; Lane, W. S.; Kopaciewicz in *Proceedings 48th ASMS Conference on Mass spectrum. Allied Topics;* Long Beach, Ca, June 11-15, **2000**, MPE 193.

(16) Dai, Y.; Randy M. Whittal, R.M; Li, L. *Anal. Chem.* **1999**, 71, 1087 –1091.

(17) Whittal, R. M.; Li, L. *Anal. Chem.* **1995**, 67, 1950-1954.

(18) Fox, P. F.; Singh, T. K.; McSweeney, P. L. H. In Chemistry of Structure-Function Relationships in Cheese; Malin, E. L., Tunick, M. H. Eds.; Plenum Press: New York, **1995**, *367*, 69-98.

Chapter 7

Conclusions and Future Work

This thesis has focused on presenting novel methods that improve both the success rate as well as the general applicability of MS-based techniques for protein identification and characterization. These methods embody two main aspects of protein sample preparation, namely techniques for enzymatic digestion of proteins prior to MS analysis, and techniques for sample cleanup, separation and enrichment. The described methods are directed at MALDI mass spectrometry for ionization and mass analysis of protein samples.

In Chapter 2, a MALDI sample deposition method was applied to the analysis of protein samples containing the surfactant, sodium dodecyl sulfate. SDS is commonly employed in protein sample workup procedures, such as SDS-PAGE gel electrophoresis and in the extraction and solubilization of hydrophobic protein from biological sources. However, the MS analysis of samples containing even low concentrations of SDS has proven to be difficult; SDS has been considered to be detrimental in the ionization of proteins with the MALDI technique. The results described in Chapter 2 demonstrate the utility of a two-layer deposition method for MALDI sample preparation for the analysis of SDS-containing proteins. It was found that strong protein signals could be achieved from samples containing up to 5% SDS by careful control of the MALDI sample deposition protocol. Several factors associated with MALDI sample preparation must be carefully controlled in order to achieve optimal performance with the MS analysis.

Chapters 3 and 4 present methods for the generation of peptide mass maps from dilute, contaminated protein solutions. The procedures presented in these chapters involve the adsorption of protein samples onto a hydrophobic, chromatographic support, followed by tryptic digestion of the adsorbed protein. By way of adsorption onto the support, the protein samples are concentrated from the dilute solution, which offers an improvement in digestion kinetics, as well as in MS sensitivity. Also, contaminants such as non-volatile salts are removed by incorporating a washing step in this procedure, allowing for effective MS analysis of highly contaminated samples. Fundamental aspects related to the digestion of adsorbed proteins were further explored in Chapter 4. It was found that while the chemical composition of the support had little effect on the digestion process, the adsorption of proteins to a surface results in partial digestion at each possible digestion site of the protein. Ultimately this partial digestion leads to the generation of a large number of peptide fragments. This can ultimately lead to an improvement in the sequence coverage from the MALDI analysis of digested protein samples.

In Chapter 5, a dual enzyme digestion protocol was presented for improving the confidence of protein identification with computer database searching techniques. Although peptide mass mapping is an effective technique for protein identification, the data generated by MALDI analysis of tryptic digests often generates a multiple list of possible sequence matches for the unknown sample. A simple dual enzyme digestion protocol involving trypsin digestion followed by N-terminal cleavage with leucine aminopeptidase results in the generation of short amino acid sequence tags. These short tags therefore provide increased information available for protein identification, thus increasing the possibility of generating unique matches for the unknown sample. The application of the dual enzyme digestion technique is extended to MS/MS fragmentation of peptides in a MALDI Q-TOF system. High quality MS/MS spectra were obtained from N-terminally cleaved peptides, which can be used to increase the sequence information content of peptides.

Finally, MS analysis of post-translation modifications of proteins was demonstrated in Chapter 6. One of the main problems associated with MS analysis of one such modification, namely phosphorylation of proteins, is related to the low stoichiometric abundance to which proteins are phosphorylated in a biological system. The work presented in Chapter 6 describes an initial study into the applicability of the IMAC technique for enrichment of phosphoproteins prior to MS analysis. The IMAC technique was first shown to effectively enrich phosphopeptides from a mixture of non-phosphorylated peptides. However, binding of acidic peptides to the IMAC column resulted in enrichment of some non-phosphorylated components as well. When intact proteins were subjected to IMAC enrichment, significant non-specific binding of the non-phosphorylated component occurred. However, selective elution of the standard phosphoprotein from the IMAC column with an aqueous solution at high pH resulted in almost complete separation of α-casein from cytochrome c.

Identification and further characterization of proteins by mass spectrometry is a relatively new technique in the history of biopolymer analysis. However, the speed, sensitivity, and widespread applicability of MS-based techniques in protein analysis have already established this instrument as one of the primary tools for this application. The methods described in this thesis further extend the applicability of mass spectrometry for protein identification and characterization.

Currently, the analysis of proteins has evolved into a much more demanding field known as proteomics. The field of proteomics is based on a high-throughput format, attempting to characterize the complete protein content of a cell or tissue culture in a particular state. An area of future work associated with this thesis is therefore devoted to the extension of the methods described in Chapters 3 and 4 for high-throughput characterization of protein samples by mass spectrometry.

A potential method, which further extends on the microcolumn digestion procedure described in Chapter 4, is currently being developed. This method involves the capture and digestion of proteins in microcolumns in an automated fashion. In this procedure, a computer-controlled HPLC system is used to manipulate protein samples, as well as washing, buffering, and enzyme solutions through a microcolumn formed by packing small ID tubing with chromatographic resin. The detection system can be directly linked for on-line MS analysis. An ion trap mass spectrometer, having MS/MS capabilities, with electrospray ionization is used to analyze the flow stream of digested protein samples from the microcolumn. Initial work on such a system has demonstrated that a standard protein sample can be adsorbed, digested, and analyzed by ESI-MS with an ion trap using the Chemstation software program to control an Agilent 1100 series HPLC system that manipulates the sample in the microcolumn.

A second approach for high-throughput automated digestion and characterization of proteins based on a MALDI-MS platform is also being explored. In this protocol, a MALDI sample plate - sample well system has been engineered which would allow for parallel capture, cleanup, and digestion of protein samples in a fashion similar to the method described in Chapter 3. An image of the prototype sample plate and well system is shown in Figure 7.1.

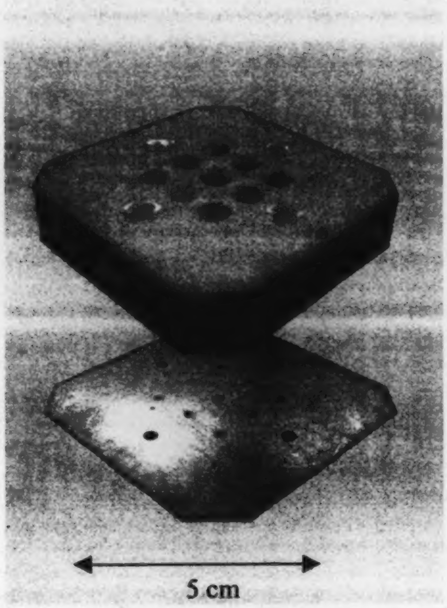

Figure 7.1 Image of sample well and frit plate.

The system is designed with two main components. The first is a stainless steel plate containing porous frits arranged in an array, which forms the bottom of the sample wells. The second component completes the sample wells by attaching a removable ~1 cm thick Teflon plate with wells arranged in the same array onto the stainless steel frit-plate. The frits will retain micro-beads added to sample solutions in the wells, and allow for the solution itself to be drawn through the porous frit with a vacuum system placed below the frit-plate. The frit-plate also acts as the MALDI sample target following sample manipulation in the wells. Initial results have been obtained with this system, including the rapid digestion and analysis of dilute protein solutions. Contaminated solutions consisting of 100-nM protein can be routinely captured, washed, digested on Poros R2 beads, and subsequently analyzed by MALDI-TOF-MS using the system. The entire sample preparation procedure can be completed in as little as 10 to 15 minutes, including only 5 minutes for rapid protein digestion. The system has the potential to be automated with an automated sample manipulation robotic system.

It is hoped that the automation and further application of the methods described in this thesis will provide useful techniques in the field of proteomics. Methods are continuously being developed to extend the applicability of the MS technique for protein characterization. Based on the incredible demands imposed by the field of proteomics, it is conceived that the development of such protocols for protein characterization will continue to be an area of considerable research for several years to come.